HAL open science

Applications des processus de Lévy et processus de branchement à des études motivées par l'informatique et la biologie

Vincent Bansaye

▶ To cite this version:

Vincent Bansaye. Applications des processus de Lévy et processus de branchement à des études motivées par l'informatique et la biologie. Mathematics [math]. Université Pierre et Marie Curie - Paris VI, 2008. English. NNT : . tel-00339230

HAL Id: tel-00339230
https://theses.hal.science/tel-00339230

Submitted on 17 Nov 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Pierre et Marie Curie

THÈSE

présentée pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ PIERRE ET MARIE CURIE

Spécialité : **Mathématiques**

soutenue par

Vincent Bansaye

Applications des processus de Lévy et processus de branchement à des études motivées par l'informatique et la biologie.

Directeurs de thèse : Jean **Bertoin** et Amaury **Lambert**

Rapporteurs :	M.	Philippe	CHASSAING	Université Henri Poincaré.
	M.	Vladimir	VATUTIN	Sketlov Mathematical Institute.

Soutenue publiquement le **14 novembre 2008** devant le jury composé de

M. Andrew	**BARBOUR**	Universität Zürich-Irchel	Examinateur
M. Jean	**BERTOIN**	Université Pierre et Marie Curie	Directeur
M. Philippe	**CHASSAING**	Université Henri Poincaré	Rapporteur
M. Jean-François	**DELMAS**	Ecole Nationale des Ponts et Chaussées	Examinateur
M. Amaury	**LAMBERT**	Université Pierre et Marie Curie	Directeur
Mme. Sylvie	**MELEARD**	Ecole Polytechnique	Examinatrice

Remerciements

Tout d'abord, je tiens ici à remercier chaleureusement mes deux directeurs de thèse, Jean Bertoin et Amaury Lambert, qui m 'ont parfaitement guidé et beaucoup appris pendant cette thèse. Si ces trois dernières années ont été vraiment agréables et intéressantes pour moi, tant scientifiquement que humainement, et m'ont donné envie de faire la recherche, il est clair que c'est beaucoup grâce a eux.

Je suis à la fois honoré et heureux que Philippe Chassaing et Vladimir Vatutin aient accépté de rapporter cette thèse et de la présence dans le jury de Andrew Barbour, Jean Francois Delmas et Sylvie Méléard. C'est aussi pour moi l'occasion d'exprimer ma reconnaissance à ces personnes qui m'ont, à différents titres, beaucoup apporté pendant cette thèse.

Mes remerciements vont aussi naturellement à mes amis. Je pense bien sur à Boris et Claire, Chi, Nicolas et Sébastien qui ont beaucoup compté pour moi ces dernières années, dans les bons moments et les moins bons. Une dédicace spéciale à Alain, Bertrand, Mathieu et la famille Casaert dont l'amitié s'est souvent transformée en une aide précieuse et un accueil chaleureux, ainsi qu' à Abbas, Etienne, François, Nathalie et Renaud. A ces remerciements j'associe mes collègues et amis de Chevaleret, parmi lesquelles Julien B., Julien S., Fernando, Mohamed, Paul, Sophie et Thanh Mai, qui ont été aussi de joyeux compagnons de bureau, de pause café, de cantine, de foot ou de TDs. Ainsi que les membres de l'administration : Isabelle, Valérie, Maria, Jacques, Philippe et Josette, dont j'ai pu apprécier depuis mon arrivée à Chevaleret l'efficacité et la gentillesse.

Je profite aussi de l'occasion pour remercier les membres de la famille Bansaye au complet. Parce que tout le monde le fait et qu'ils m'en auraient voulu sinon. Mais aussi parce que depuis le début étudier a été facile et agréable avec eux, qu'ils m'ont toujours formidablement soutenu, et que cette page est l'occasion de leur témoigner ma reconnaissance.

Je tiens enfin à saluer les personnes avec qui j'ai travaillé ou partagé les temps libres des congrès, et notamment mes grands frères et grandes soeurs de thèse. J'ai passé avec elles de très bons moments qui m'ont donné envie de participer aux rencontres ou groupes de travail, de collaborer, de découvrir les interfaces avec d'autres disciplines...

Ma dernière pensée va à Gilbert Lemaitre, qui en plus d'avoir été un professeur de mathématiques dynamique et inventif au collège pendant de nombreuses années et un joueur de badminton redoutable, était un bon vivant et un esprit curieux qui nous manque...

Contents

Chapter 1

Introduction et présentation des résultats

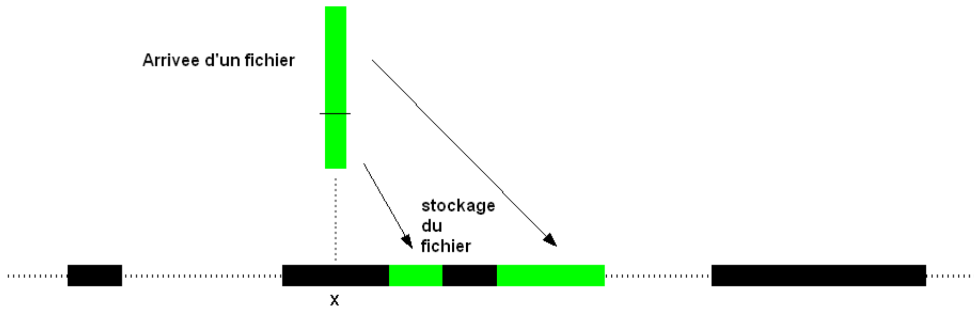
Ce document rassemble les résultats obtenus pendant trois années de thèse sous la direction de Jean Bertoin et Amaury Lambert. Il se décompose en deux parties.

La première partie traite d'un modèle stochastique de stockage de données en temps continu, dont la théorie a été développée principalement pour ses applications à l'informatique [29, 81]. Elle est principalement constituée des deux articles suivants :

- (2007). On a model for the storage of files on a hardware I : Statistics at a fixed time and asymptotics. *Prépublication.* Soumis.
- (2007). On a model for the storage of files on a hardware II : Evolution of a typical data bloc. *J. Appl. Prob.* Vol 44, no 4, 901-927.

Dans ce modèle, le disque dur est représenté par la droite réelle. Les temps d'arrivée des fichiers, les emplacements où l'on cherche à les stocker et leur taille sont aléatoires. Les hypothèses naturelles d'indépendance et de stationnarité pour ces quantités conduisent à modéliser ces arrivées de fichiers par un processus ponctuel de Poisson [62, 73, 84]. Chaque fichier est dirigé sur un emplacement du disque et stocké le plus près possible à droite de celui-ci : Si cet emplacement est déjà occupé, on cherche le premier emplacement de libre à droite et lorsque un fichier est plus gros que l'espace libre rencontré, il est fragmenté pour être stocké en plusieurs morceaux.

Figure 1. Arrivée et stockage d'un fichier (en clair) sur le disque où les parties occupées sont en noir.



Ce modèle est une version continue du problème de parking de Knuth, qui décrit le stockage de fichiers de taille unité qui arrivent successivement sur un disque possédant n emplacements [29, 37, 41]. Knuth était initialement intéressé par le décalage que subit un fichier pour trouver un espace libre, de manière à contrôler le temps qu'il faut pour le retouver. Pour ce modèle, P. Chassaing et G. Louchard [28] ont établi une transition de phase pour la taille du plus gros bloc de données du disque lorsque $n \to \infty$, qui est décrite par le coalescent additif. Puis J. Bertoin et G. Miermont [23] ont généralisé ces résultats pour des fichiers de taille aléatoire. On considère donc ici une version continue de ce dernier modèle avec un disque de taille infinie.
Ce modèle a de nombreux liens avec les files d'attentes M/G/1 (voir par exemple [30, 81, 82]). En effet, si l'on voit les emplacements où l'on veut stocker les fichiers comme les instants d'arrivées de clients d'une file d'attente et la taille du fichier comme le temps de service du client, l'espace occupé par les données à temps fixe devient la période de service d'une file d'attente M/G/1.

Tout d'abord, nous avons caractérisé géométriquement et analytiquement l'espace occupé par les données à un temps fixe et considéré certains exemples. Nous pouvons alors donner la loi du décalage subi par un fichier pour commencer son stockage, ainsi que le dernier point utilisé pour l'enregister.
En utilisant ces résultats et en s'inspirant de [28], nous avons décrit le comportement asymptotique de l'espace occupé par les données près du temps de saturation du disque ainsi que la taille du plus gros bloc de données. On retrouve la transition de phase observée par P. Chassaing et G. Louchard et les théorèmes limites dépendent de la queue de distribution de la taille des fichiers, comme dans l'article de J. Bertoin et G. Miermont.
Nous avons ensuite réalisé une étude dynamique du processus de stockage, en caractérisant l'évolution en temps d'un bloc de données typique. Pour cela, nous nous appuyons à nouveau sur les caractérisations géométriques et analytiques à temps fixe. On démontre ainsi que les instants de sauts de l'extrémité gauche du bloc de donnée typique s'accumulent suivant la 'stick breaking sequence' de Pitman [80]. De plus les quantités successives de données arrivées à gauche du bloc et stockées à sa droite forment une suite iid. Enfin, la longueur du bloc suit un processus de branchement en temps continu avec immigration inhomogène.

Le fait de modéliser par un processus ponctuel de Poisson les arrivées de fichiers permet d'utiliser abondamment la théorie des processus de Lévy et de ses fluctuations [19, 86] ainsi que les ensembles régénératifs [71, 89, 90], tant pour la construction du modèle que les comportements asymptotiques et l'étude dynamique.

La seconde partie concerne des questions liées à la biologie, pour lesquelles on utilise des processus de branchement (voir e.g. [48, 60, 65]). Plus précisément, nous nous sommes intéressé à la prolifération de parasites dans des cellules en division et au processus de branchement en environnement aléatoire. Cette partie est principalement formée de 3 articles :

- (2007). Proliferating parasites in dividing cells : Kimmel's branching model revisited. *Ann. Appl. Probab.*, Vol 18, no 3, 967-996.
- (2008). Surviving particles for subcritical branching processes in random environment. Révisé pour *Stoch. Proc. Appl.*
- (2008). Cell contamination and branching processes in random environment with immigration. *Prépublication.* Soumis.

M. de Paepe, G. Paul and F. Taddei du Laboratoire TaMaRa (Hôpital Necker, Paris) [88] ont observé pour la bacterie E-Coli une répartition très inégale des parasites de la bactérie dans ses deux cellules filles au moment de la division. Cette asymétrie est surprenante puisqu'on s'attend à un partage équitable du contenu biologique d'une cellule. Nous avons considéré un modèle en temps discret pour rendre compte de la multiplication aléatoire d'un parasite avec une répartition aléatoire (éventuellement très inéquitable) des parasites au moment de la division. Pour cela nous nous sommes inspirés du modèle de M. Kimmel [59] en temps continu avec répartition symétrique de parasites au moment de la division. Ainsi, pour des raisons pratiques, nous distinguons une première cellule fille notée 0 et une deuxième notée 1 et nous condensons multiplication d'un parasite et répartition de sa descendance en une multiplication à deux types. Chaque parasite se comporte alors indépendamment et donne naissance à chaque génération à $Z^{(0)}+Z^{(1)}$ parasites, parmi lesquels $Z^{(0)}$ vont dans la première cellule fille et $Z^{(1)}$ dans la deuxième au moment de la division. On autorise dissymétrie et dépendance pour le couple de variable aléatoire $(Z^{(0)}, Z^{(1)})$.
A.D. Barbour, M. Kafetzaki, C.J. Luchsinger et M.J. Luczak [16, 17, 67, 68] ont une approche différente pour modéliser l'infection des cellules. Ils considèrent un processus de branchement à une infinité de types où le type donne le nombre de parasites de la cellule. Ils démontrent que la limite en grande population de cellules infectées est décrite par une équation différentielle déterministe.
Par ailleurs, Julien Guyon [47] a modélisé les asymétries au moment de la division

pour le problème du vieillissement cellulaire et obtenu des résultats généraux pour des cellules en division en considérant des arbres markoviens dissymétriques. Mais ses résultats ne s'appliquent pas à notre problème, puisque dans notre cas le nombre de parasites dans une lignée cellulaire typique a un comportement asymptotique dégéneré (i.e. il tend vers 0 ou ∞ p.s.).

Tout d'abord, nous avons déterminé sous quelles conditions l'organisme guérit, au sens où la proportion de cellules infectées tend vers zéro. Ceci met en évidence l'intérêt d'une répartition inéquitable des parasites pour la cellule. Nous avons ensuite établi le comportement asymptotique du nombre de cellules infectées et des proportions de cellules infectées par un nombre donné de parasite. Les résultats dépendent du couple de valeurs $(\mathbb{E}(Z^{(0)}), \mathbb{E}(Z^{(1)}))$ et donnent des asymptotiques déterministes.

Pour obtenir ces résultats, nous utilisons deux processus de branchement classiques associés à ce modèle.
Premièrement, le nombre total de parasites d'une génération suit un processus de Galton-Watson, c'est-à- dire que chaque parasite se reproduit indépendamment et avec la même loi à chaque génération. Ce processus a été beaucoup étudié [5, 8, 70] et nous utilisons un certain nombre de résultats classiques tels que le critère d'extinction, le théorème de Kesten-Stigum et la limite quasi-stationaire de Yaglom.
Deuxièmement, nous utilisons le nombre de parasites dans une lignée cellulaire aléatoire, qui suit un processus de branchement en environnement aléatoire (BPRE) [6, 7, 8, 87]. L'environnement aléatoire vient du choix de la cellule fille au moment de la division pour construire la lignée aléatoire, car il détermine les lois de reproduction successives dans la lignée. On utilise également le critère d'extinction et la limite de Yaglom. Mais un BPRE (processus de branchement en environnement aléatoire) sous-critique (i.e. qui s'éteint p.s.) peut avoir des comportements très différents. Par exemple, sa moyenne peut tendre géométriquement vers 0, être constante ou tendre géométriquement vers l'infini. De plus, l'équivalent de la probabilité de survie au temps n dépend du fait que le processus est fortement sous critique, moyennement sous critique ou faiblement sous critique [43, 46].
De manière informelle, les BPRE fortement sous critiques ont les propriétés attendues, i.e. les analogues de ceux d'un processus de Galton Watson. Tandis que les BPRE faiblement sous critiques présentent de nombreuses différences, par exemple pour le processus réduit [38]. De plus, ils vérifient des 'propriétés surcritiques', tels que l'analogue du théorème de Kesten-Stigum, comme l'on mis en évidence V. I. Afanasyev, J. Geiger, G. Kersting and V. A. Vatutin [2]. Ceci explique que les théorèmes limites pour notre modèle diffèrent en fonction des sous domaines déterminant si la lignée typique est faiblement, moyennement ou fortement sous critique.

Le domaine où les résultats établis sont partiels est naturellement le domaine correspondant au BPRE faiblement sous critique. En fait, obtenir les convergences en probabilité des proportions de cellules infectées par un nombre donné de parasites nécessite de savoir si la limite quasistationaire d'un BPRE dépend ou non du nombre de particules initiales. Cette question semble difficile dans le cas faiblement sous critique, alors qu'il est facile de prouver qu'il n'y a pas de dépendance pour un processus de Galton Watson et plus généralement pour un BPRE fortemenent sous critique. Cette question conduit plus généralement à se demander comment les théorèmes limites d'un BPRE sous critique dépendent du nombre de particules initiales.

Nous avons alors considéré un BPRE quelconque. C'est à dire que pour chaque génération, on tire de manière iid une loi de reproduction et toutes les particules de cette génération se reproduisent de manière indépendante suivant cette loi. On peut par exemple se représenter une population de fleur qui est soumis à chaque génération à un climat aléatoire. Ce climat détermine la loi de reproduction des fleurs, qui se reproduisent alors indépendamment. Nous avons prouvé que pour un BPRE faiblement sous critique, la probabilité de survie d'une population n'était pas asymptotiquement proportionnelle à la taille de la population initiale, contrairement au cas d'un BPRE fortement ou moyennement sous critique, et établi d'autres asymptotiques. De plus nous démontrons que, dans le cas faiblement sous critique, conditionnellement à la survie de la population, plusieurs particules initiales voient leur descendance survivre avec probabilité positive, ce qui ne se produit pas pour un processus de Galton Watson sous critique. Nous donnons également une interprétation au sens des environnements de ces résultats. De manière informelle, nous prouvons que la survie d'une particule en environnement fortement ou moyennement sous critique est due à l'aléa de la reproduction (la particule survit malgré l'environnement). Tandis que pour un environnement faiblement sous critique, la survie de la particule est due à l'aléa environemental (environnement particulièrement favorable pour la particule). Enfin, nous montrons que dans le cas fortement sous critique, le nombre d'individu en vie sachant que la population survit dans le futur (qui est donné par le Q-processus) converge en distribution, tandis qu'il diverge dans les deux autres cas.

Nous avons ensuite pris en compte une contamination aléatoire des cellules par des parasites extérieurs à la population. Dans ce modèle, pour des raisons biologiques et techniques, la loi du nombre de parasites qui contaminent une cellule donnée dépend (uniquement) du fait que cette cellule est déjà infectée ou non.

Le nombre de parasites dans une lignée cellulaire est alors un processus de branchement en environnement aléatoire avec immigration, où l'immigration dépend de l'état à travers le fait que celui-ci est nul ou non. Les processus de Galton Watson avec immigration sont bien connus [5, 8, 70] : si le processus est

sous critique et que l'espérance du logarithme de l'immigration est finie, le nombre de particules tend en distribution vers une variable aléatoire finie. Sinon il tend en probabilité vers l'infini. E.S. Key [58] a obtenu la convergence en distribution pour les BPRE avec immigration sous l'hypothèse d'existence de l'espérance du logarithme de l'immigration. Il considère le cadre plus général des BPRE multi-types avec immigration et donne également une estimation de la queue du temps de retour en zéro. A. Roitershtein [83] a complété ces résultats par une loi forte des grands nombres et un théorème central limite pour la somme partielle associée à ce processus.
Nous établissons les convergences pour le cas d'une immigration dépendant de l'état (zéro ou strictement positive), et traitons le cas où le logarithme de l'espérance de l'immigration est infinie. Nous fournissons également une estimation de la vitesse de convergence dépendant du nombre de particules initiales. Pour cela, nous utilisons des arguments de couplage pour nous ramener à un BPRE avec immigration et appliquer [58], ainsi que le théorème de renouvellement et sa vitesse de convergence.

Ces résultats donnent le comportement du processus comptant le nombre de parasites dans une lignée cellulaire en présence d'une contamination extérieure. En utilisant [47], on obtient alors une loi des grands nombres sur les proportions asymptotiques de cellules infectées par un nombre donné de parasites. En effet, l'immigration empêche l'absorption (le processus ne tend pas vers zéro p.s.), ce qui permet d'appliquer les théorèmes généraux sur les chaines de Markov indexées par un arbre. A nouveau, nous constatons l'interêt d'une division assymétrique pour l'organisme.

Résultats principaux de la partie I

Les fichiers sont étiquetés par $i \in \mathbb{N}$ et l'on note $t_i \geq 0$ le temps d'arrivée du fichier i, $x_i \in \mathbb{R}$ l'emplacement où l'on souhaite stocker son extrémité gauche et $l_i \geq 0$ sa taille.
L'ensemble $\{(t_i, x_i, l_i) : i \in \mathbb{N}\}$ est un Processus Ponctuel de Poisson (PPP) d'intensité $\mathrm{d}t \otimes \mathrm{d}x \otimes \nu$ et l'on note

$$\mathrm{m} := \int_0^\infty l\nu(\mathrm{d}l) < \infty$$

la quantité moyenne de données qui arrivent dans un intervalle de temps unité sur une unité du disque. On commence par construire ce modèle et remarquer que le disque est plein au temps $1/\mathrm{m}$.

On utilise pour cela le processus $(Y_x^{(t)})_{x\in\mathbb{R}}$ défini par

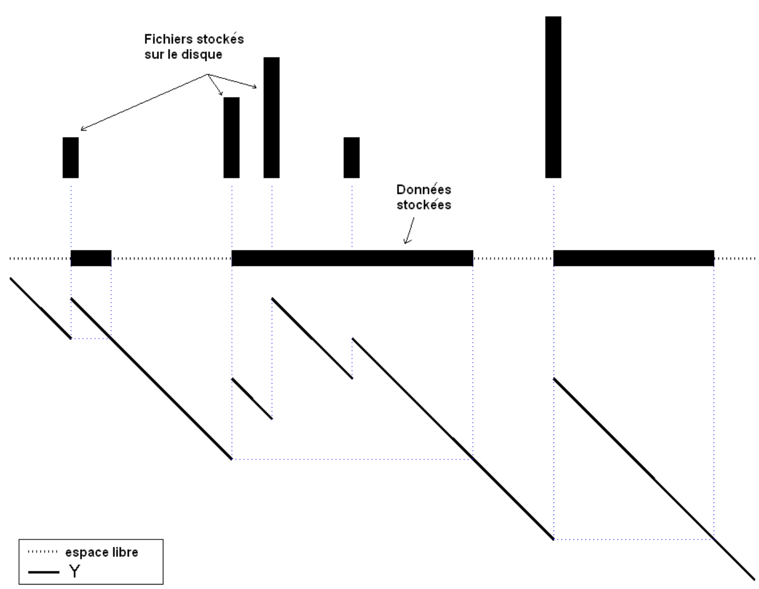
$$Y_0^{(t)} := 0 \quad ; \quad Y_b^{(t)} - Y_a^{(t)} = -(b-a) \ + \sum_{\substack{t_i \leq t \\ x_i \in]a,b]}} l_i \quad \text{si } a < b.$$

Ce processus est un processus à accroissement indépendants et stationnaires (processus de Lévy) à variation bornée et dérive égale à un, dont les sauts donnent les emplacements d'arrivée et les tailles des fichiers. On introduit également le processus infimum :

$$I_x^{(t)} := \inf\{Y_y^{(t)} : y \leq x\}.$$

Comme on le devine sur la figure suivante, la partie du disque dur occupée au temps t forme un ensemble aléatoire $\mathcal{C}(t)$ égal à l'ensemble des points où $Y^{(t)}$ est au-dessus de son infimum passé :

Figure 2. Stockage des fichiers et processus de Lévy associé.



Le résultat peut s'énoncer ainsi

Propriété. *Pour tout* $t < 1/\mathrm{m}$*,* $\mathcal{C}(t) = \{x \in \mathbb{R} : \ Y_x^{(t)} > I_x^{(t)}\} \neq \mathbb{R}$ *p.s.*
Pour tout $t \geq 1/\mathrm{m}$*,* $\mathcal{C}(t) = \mathbb{R}$ *p.s.*

Nous pouvons alors établir quelques propriétés géométriques de l'ensemble occupé au temps t, en décrivant son complémentaire $\mathcal{R}(t)$ qui est égal à l'espace libre au temps t.

Propriété. *Pour tout* $t \geq 0$*,* $\mathcal{R}(t)$ *est stationnaire, sa fermeture est symétrique par rapport* 0 *en distribution et forme un ensemble régénératif.*
De plus, pour tout $x \in \mathbb{R}$*,* $\mathbb{P}(x \in \mathcal{C}(t)) = \min(1, \mathrm{m}t)$*.*

Nous donnons ensuite une description analytique de l'espace libre en distinguant le bloc de données au temps t qui contient le point zéro, que l'on note $\mathbf{B}_0(t)$. On introduit également l'extrémité de gauche (resp. droite) $g(t)$ (resp. $d(t)$) de ce bloc :

$$g(t) = \sup\{y \leq 0 : y \in \mathcal{R}(t)\}, \qquad d(t) = \inf\{y > 0 : y \in \mathcal{R}(t)\}.$$

$$\mathbf{B}_0(t) = [g(t), d(t)].$$

Les parties libres à gauche et à doite de $\mathbf{B}_0(t)$ sont deux ensembles régénératifs de $[0, \infty)$ indépendants identiquements distribués, qui sont indépendants de $\mathbf{B}_0(t)$. De plus, ces deux ensembles régénératifs sont les images de subordinateurs indépendants dont l'exposant de Laplace est la fonction réciproque de l'exposant de Laplace de $(Y_x^{(t)})_{x\in\mathbb{R}}$. Pour achever la description à temps fixe, nous donnons la loi de $\mathbf{B}_0(t)$.

Régimes asymptotiques.

Nous établissons ici le comportement asymptotique du disque près de l'instant de saturation. Nous donnons tout d'abord quelques définitions utiles pour énoncer les résulats. Reprenant les notations de [23], nous notons $\nu \in \mathcal{D}_{2+}$ si le moment d'ordre 2 de ν est fini et $\mathrm{m}_2 := \int_0^\infty l^2 \nu(\mathrm{d}l)$. Pour tout $\alpha \in]1, 2]$, nous notons $\nu \in \mathcal{D}_\alpha$ si

$$\exists C > 0 \ \text{ tel que } \ \nu[x, \infty) \overset{x\to\infty}{\sim} C x^{-\alpha}.$$

Nous posons alors, pour tout $\alpha \in]1, 2[$:

$$C_\alpha := \left(\frac{C\Gamma(2-\alpha)}{\mathrm{m}_2(\alpha-1)}\right)^{1/\alpha}.$$

Nous introduisons également le mouvement brownien indexé par $\mathbb{R}$ noté $(B_z)_{z\in\mathbb{R}}$, i.e. $(B_x)_{x\geq 0}$ et $(B_{-x})_{x\geq 0}$ sont deux mouvements browniens indépendants. Pour tout $\alpha \in]1, 2[$, nous définissons $(\sigma_z^{(\alpha)})_{z\in\mathbb{R}}$ le processus càdlàg à accroissements indépendants et stationnaires tel que $(\sigma_x^{(\alpha)})_{x\geq 0}$ est un processus de Lévy stable d'index α sans sauts négatifs :

$$\forall\ x \geq 0, \lambda \geq 0, \quad \mathbb{E}\big(\exp(-\lambda\sigma_x^{(\alpha)})\big) = \exp(x\lambda^\alpha).$$

Enfin, nous introduisons les processus suivants indéxés par $\mathbb{R}$ pour tout $\lambda \geq 0$ et $\alpha \in]1, 2[$,

$$Y_z^{2+,\lambda} = -\lambda z + \sqrt{\mathrm{m}_2/\mathrm{m}}B_z, \qquad Y_z^{2,\lambda} = -\lambda z + \sqrt{C/\mathrm{m}}B_z, \qquad Y_z^{\alpha,\lambda} = -\lambda z + C_\alpha \sigma_z^{(\alpha)},$$

et le processus infimum associé défini pour $x \in \mathbb{R}$ par $I_x^{\alpha,\lambda} := \inf\{Y_y^{\alpha,\lambda} : y \leq x\}$.

En utilisant les fonctions suivantes définies pour tout $t \in [0, 1/\mathrm{m}[$ et $\alpha \in]1, 2[$ par

$$\epsilon_{2+}(t) = (1 - \mathrm{m}t)^2, \qquad \epsilon_2(t) = 2\frac{(1 - \mathrm{m}t)^2}{-\log((1 - \mathrm{m}t))}, \qquad \epsilon_\alpha(t) = (1 - \mathrm{m}t)^{\frac{\alpha}{\alpha-1}},$$

nous obtenons

Théorème. *Si $\nu \in \mathcal{D}_\alpha$ ($\alpha \in]1, 2] \cup \{2+\}$), alors $\epsilon_\alpha(t).\mathcal{R}(t)^{cl}$ converge en distribution pour la distance de Hausdorff vers $\{x \in \mathbb{R} : Y_x^{\alpha,1} = I_x^{\alpha,1}\}^{cl}$ quand $t \to 1/\mathrm{m}$.*

Ce qui implique

Corollaire. *Si $\nu \in \mathcal{D}_\alpha$ ($\alpha \in]1, 2] \cup \{2+\}$), alors $\epsilon_\alpha(t).(g(t), d(t))$ converge en distribution quand $t \to 1/\mathrm{m}$ vers $(\sup\{x \leq 0 : Y_x^{\alpha,1} = I_0^{\alpha,1}\}, \inf\{x \geq 0 : Y_x^{\alpha,1} = I_0^{\alpha,1}\})$. Si $\nu \in \mathcal{D}_{2+}$ (resp. $\mathcal{D}_2$), $\epsilon_\alpha(t).(d(t) - g(t))$ converge en distribution vers une loi gamma de paramètre $(1/2, \mathrm{m}/(4\mathrm{m}_2))$ (resp. $(1/2, \mathrm{m}/4)$).*

Nous considérons maintenant le disque restreint à un segment $[0, x]$ et introduisons les fonctions définies pour $x \geq 1$ et $\alpha \in]1, 2[$ par

$$f_{2+}(x) = 1/\sqrt{x}, \qquad f_2(x) = \sqrt{\log(x)/x}, \qquad f_\alpha(x) = x^{1/\alpha - 1}.$$

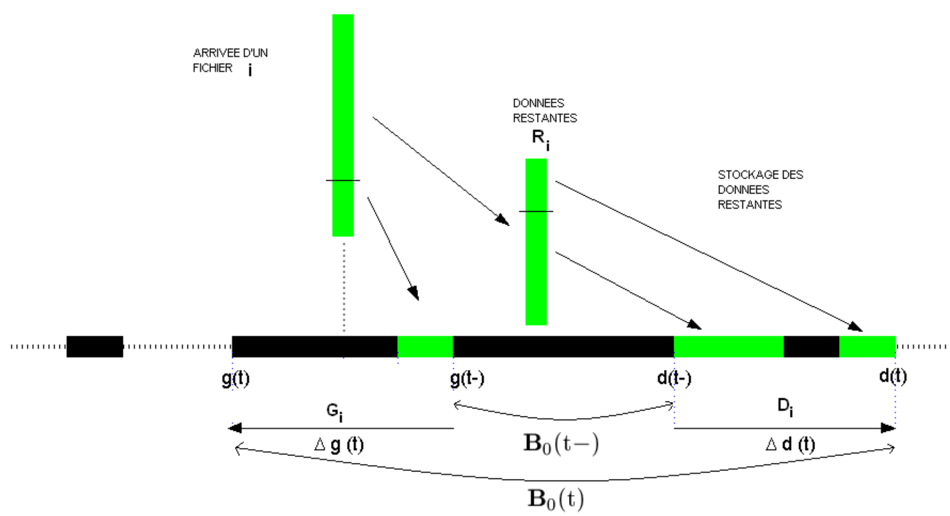
Théorème. *Si $\nu \in \mathcal{D}_\alpha$ ($\alpha \in]1, 2] \cup \{2+\}$), $x \to \infty$ et $t \to 1/\mathrm{m}$ avec $1 - \mathrm{m}t \sim \lambda f_\alpha(x)$ et $\lambda > 0$, alors $x^{-1}(\mathcal{R}(t)^{cl} \cap [0, x])$ converge faiblement pour la distance de Hausdorff vers $\{z \in [0, 1] : Y_z^{\alpha,\lambda} = I_z^{\alpha,\lambda}\}^{cl}$.*

Comme P. Chassaing et G. Louchard [28], on observe ainsi une transition de phase pour la taille du plus gros bloc de données sur $[0, x]$ au temps t, noté $B_1(x, t)$ quand $x \to \infty$ suivant le taux de remplissage du disque.

Corollaire. *Soit* $\nu \in \mathcal{D}_\alpha$ *(*$\alpha \in]1,2] \cup \{2+\}$*),* $x \to \infty$ *et* $t \to 1/\mathrm{m}$*:*
- Si $1 - \mathrm{m}t \sim \lambda f_\alpha(x)$ *avec* $\lambda > 0$*, alors* $B_1(x,t)/x$ *converge en distribution vers la plus grande excursion de* $(Y_x^{\alpha,\lambda} - I_x^{\alpha,\lambda})_{x\in[0,1]}$.
- Si $f_\alpha(x) = o(1 - \mathrm{m}t)$*, alors* $B_1(x,t)/x \xrightarrow{\mathbb{P}} 0$.
- Si $1 - \mathrm{m}t = o(f_\alpha(x))$*, alors* $B_1(x,t)/x \xrightarrow{\mathbb{P}} 1$.

Evolution d'un bloc de données typique. Nous réalisons maintenant une étude dynamique du disque en caractérisant l'évolution en temps du bloc de données en 0, $\mathbf{B}_0$. Ce bloc peut croitre pour deux types d'événements différents

- Un fichier $i \in \mathbb{N}$ arrive à gauche de $\mathbf{B}_0$ au temps T_i et il ne peut être entièrement stocké à sa gauche. Il provoque alors un accroissement G_i de l'extrémité gauche du bloc. Les données du fichier R_i qui ne peuvent être stockées à gauche sont appelées *données restantes.* Elles provoquent un accroissement D_i de l'extrémité droite.



- Un fichier $i \in \mathbb{N}$ arrive sur $\mathbf{B}_0$ et provoque alors (uniquement) un saut de de l'extremité droite.

En premier lieu, nous nous intéressons à l'extremité gauche et l'on peut noter que

$$g(t) := -\sum_{T_i \leq t} G_i.$$

Théorème. *Les temps de saut de $(g(t))_{t\in[0,1/\mathrm{m}]}$ forment une suite croissante $(T_i)_{i\in\mathbb{N}}$ qui s'accumule en $1/\mathrm{m}$, dont la distribution ne dépend pas de ν. Plus précisement, en posant $T_0 = 0$, pour tout $i \geq 1$, conditionnellement à $T_{i-1} = t$, T_i est indépendant de $(T_j)_{0\leq j\leq i-1}$ et uniformément sur $[t, 1/\mathrm{m}]$.*

De plus, $\{(T_i, G_i) : i \in \mathbb{N}\}$ est un PPP sur $[0, 1/\mathrm{m}[\times\mathbb{R}^+$ d'intensité

$$\mathrm{d}t\mathrm{d}x \int_0^\infty \mathbb{P}(Y_x^{(t)} \in -\mathrm{d}l)\bar{\nu}(l).$$

En d'autre termes, $(g(t))_{t\in[0,1/\mathrm{m}]}$ est un processus additif dont le triplet générateur est égal à

$$\left(0, \int_0^t \mathrm{d}s \int_0^\infty \mathbb{P}(Y_x^{(s)} \in -\mathrm{d}l)\bar{\nu}(l), 0\right).$$

Nous prouvons que $\{(T_i, G_i, R_i) : i \in \mathbb{N}\}$ et $\{(T_i, G_i, D_i) : i \in \mathbb{N}\}$ forment des PPP dont nous donnons l'intensité. Ceci permet de donner les taux de transitions de $\mathbf{B}_0$. Nous obtenons en particulier une description des processus donnant les extrémités de $\mathbf{B}_0$ $(g(t))_{t\geq 0}$ et $(d(t))_{t\geq 0}$, ainsi que de la longueur de $\mathbf{B}_0$ $(d(t) - g(t))_{t\geq 0}$.
Pour ne pas présenter ici des résultats trop techniques, nous ne donnons pas la description précise de $(\mathbf{B}_0(t))_{t\geq 0}$ et renvoyons aux Sections 5.3 et 5.4. En revanche, il est intéressant de noter que la suite formée des quantités de données restantes est iid :

Proposition. *$\{(T_i, R_i) : i \in \mathbb{N}\}$ est un PPP sur $[0, 1/\mathrm{m}[\times\mathbb{R}^+$ d'intensité*

$$\mathrm{d}t\mathrm{d}z\frac{\bar{\nu}(z)}{1 - \mathrm{m}t}.$$

En d'autres termes, $(R_i)_{i\in\mathbb{N}}$ est une suite iid indépendante de $(T_i)_{i\in\mathbb{N}}$ dont la distribution est donnée par

$$\mathbb{P}(R_i \in \mathrm{d}z) = \mathrm{m}^{-1}\bar{\nu}(z)\mathrm{d}z, \quad z \geq 0.$$

Résultats principaux de la partie II

Processus de branchement en environnement aléatoire (BPRE) sous-critique. Le processus de branchement en environnement aléatoire est spécifié par une suite iid de fonctions génératrices $(f_n)_{n\in\mathbb{N}}$ distribuées comme une fonction génératrice aléatoire f. C'est à dire que pour tout $n \in \mathbb{N}$,

$$\mathbb{E}\big(s^{Z_{n+1}}|Z_0,\dots,Z_n;\ f_0,\dots,f_n\big) = f_n(s)^{Z_n} \qquad (0 \le s \le 1).$$

Le processus est sous critique si

$$\mathbb{E}(\log(f'(1)) < 0.$$

Dans ce cas, il s'éteint p.s. en temps fini. On distingue les BPRE fortement, moyennement et faiblement sous critique suivant que

$$\mathbb{E}(f'(1)\log(f'(1)))$$

est négative, nulle ou positive. Le processus de Galton Watson sous-critique est fortement sous-critique. Dans le chapitre 7, nous déterminons comment les théorèmes limites dépendent du nombre initial de particules. Pour cela on introduit

$$\alpha_k := \lim_{n\to\infty} \mathbb{P}_k(Z_n > 0)/\mathbb{P}_1(Z_n > 0).$$

et le réel $\alpha \in [0,1]$ caractérisé par

$$\gamma := \inf_{\theta\in[0,1]} \big\{\mathbb{E}\big(f'(1)^\theta\big)\big\} = \mathbb{E}\big(f'(1)^\alpha\big).$$

On peut noter que $\alpha \in]0,1[$ dans le cas faiblement sous critique et

Théorème. *Dans les cas fortement et moyennement sous critiques, pour tout $k \in \mathbb{N}$, $\alpha_k = k$.*

Dans le cas faiblement sous critique, $\alpha_k \to \infty$ quand $k \to \infty$ et il existe $M_+ > 0$ tel que

$$\alpha_k \le M_+ \log(k) k^\alpha, \quad (k \ge 2).$$

Si l'on suppose de plus que $\mathbb{E}(f'(1)^{1/2}\log(f'(1))) > 0$ (i.e. $\alpha < 1/2$) et que $f''(1)/f'(1)$ est borné, il existe $M_- > 0$ tel que

$$\alpha_k \ge M_- \log(k) k^\alpha, \quad (k \in \mathbb{N}).$$

Ensuite nous regardons si plusieurs particules initiales peuvent survivre conditionnellement à la survie de la population. En notant $Z_n^{(i)}$ le nombre de descendants à la génération n de la particule initiale i, nous prouvons que conditionnelement à la survie de la population en temps long, plusieurs particules initiales peuvent survivent dans le cas faiblement sous-critique.

Proposition. *Dans les cas fortement et moyennement sous critiques, pour tout* $k \geq 1$,

$$\lim_{n\to\infty} \mathbb{P}_k(\exists i \neq j,\ 1 \leq i,j \leq k,\ Z_n^{(i)} > 0,\ Z_n^{(j)} > 0\ |Z_n > 0) = 0.$$

Dans le cas faiblement sous critique, pour tout $k \geq 1$,

$$\lim_{n\to\infty} \mathbb{P}_k(\forall i,\ 1 \leq i \leq k,\ Z_n^{(i)} > 0\ |Z_n > 0) > 0.$$

Nous donnons alors le comportement asymptotique du nombre de particules initiales qui survivent au temps n, N_n, conditionnellement à la survie de la population, pour certains BPRE faiblement sous critiques :

Théorème. *Si* $\mathbb{E}(f'(1)^{1/2}\log(f'(1))) > 0$ *(i.e.* $\alpha < 1/2$*) et* $f''(1)/f'(1)$ *borné, il existe* $A_l \downarrow_{l\to\infty} 0$ *tel que pour tous* $k \geq l \geq 0$,

$$\limsup_{n\to\infty} \mathbb{P}_k(N_n \geq l \mid Z_n > 0) \leq A_l.$$

De plus, pour tout $l \in \mathbb{N}^*$,

$$\liminf_{k\to\infty} \liminf_{n\to\infty} \mathbb{P}_k(N_n = l \mid Z_n > 0) > 0.$$

Ces résultats s'expliquent en étudiant les environements selectionnés par la survie de la population. Pour cela, on introduit

$$\mathbf{f}_n := (f_0, f_1, \ldots, f_{n-1}),$$

la suite des environements jusqu'au temps n et $p(\mathbf{f}_n)$ la probabilité de survie sous cette suite d'environements

$$p(\mathbf{f}_n) := \mathbb{P}(Z_n > 0 \mid \mathbf{f}_n).$$

Théorème. *Dans les cas fortement et moyennement sous critiques, pour tous* $k \in \mathbb{N}^*$ *et* $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}_k(p(\mathbf{f}_n) \geq \epsilon \mid Z_n > 0) = 0.$$

Dans le cas faiblement sous critique, pour tous $k \geq 1$ *et* $\epsilon > 0$,

$$\liminf_{n\to\infty} \mathbb{P}_k(p(\mathbf{f}_n) \geq \epsilon \mid Z_n > 0) \xrightarrow{\epsilon\to 0+} 1.$$

<u>Processus de branchement en environement aléatoire avec immigration</u> Nous ajoutons maintenant une contamination aléatoire au BPRE. La loi de reproduction du BPRE est toujours la fonction génératrice aléatoire f. L'immigration dépend de l'état seulement à travers le fait qu'il soit nul ou non. L'immigration pour un

état nul est distribuée comme une v.a. Y_0. Pour un état non nul, elle est distribuée comme une v.a. Y_1. Plus précisement, pour tout $n \in \mathbb{N}$, conditionnellement à $Z_n = x$,

$$Z_{n+1} = Y_x^{(n)} + \sum_{i=1}^{x} X_i^{(n)},$$

avec

- $(X_i^{(n)})_{i \in \mathbb{N}}$ et $Y_x^{(n)}$ sont indépendants.
- Conditionnellement à $f_n = g$, $(X_i^{(n)})_{i \in \mathbb{N}}$ est une suite iid dont la fonction génératrice est égale à g.
- Pour tout $x \geq 1$ et $n \in \mathbb{N}$, $Y_x^{(n)} \overset{d}{=} Y_1$ et $Y_0^{(n)} \overset{d}{=} Y_0$.
- $Y_x^{(n)}$ est indépendant de $(Y_x^{(n)} : 0 \leq i \leq n-1)$.

L'étude de ce processus était initialement motivée par le modèle de division cellulaire avec contamination aléatoire de parasites décrit ensuite. Le résultat obtenu est le suivant

Théorème. *(i) Si* $\mathbb{E}\big(\log(f'(1))\big) < 0$ *et* $\max(\mathbb{E}(\log^+(Y_i)) : i = 0, 1) < \infty$*, alors il existe une variable aléatoire* Z_∞ *telle que pour tout* $k \in \mathbb{N}$*,* Z_n *partant de* k *converge en distribution vers* Z_∞ *quand* $n \to \infty$*.*

Si de plus, il existe $q > 0$ *tel que* $\max(\mathbb{E}(Y_i^q) : i = 0, 1) < \infty$*, alors pour tout* $\epsilon > 0$*, il existe* $0 < r < 1$ *et* $C > 0$ *tels que,*

$$\sum_{l=0}^{\infty} |\mathbb{P}_k(Z_n = l) - \mathbb{P}(Z_\infty = l)| \leq C k^\epsilon r^n, \quad (n, k \in \mathbb{N}).$$

(ii) Si $\mathbb{E}\big(\log(f'(1))\big) \geq 0$ *ou* $\max(\mathbb{E}(\log^+(Y_i)) : i = 0, 1) = \infty$*, alors* Z_n *tend en probabilité vers l'infini quand* $n \to \infty$*.*

Prolifération de parasites dans une cellule en division Les cellules en division forment un arbre binaire $\mathbb{T} = \cup_{i=0}^{n} \{0, 1\}^n$. On note $\mathbb{G}_n$ (resp. $\mathbb{G}_n^*$) l'ensemble des cellules (resp. cellules infectées) à la génération n et par $Z_{\mathbf{i}}$ le nombre de parasites contenus dans la cellule $\mathbf{i} \in \mathbb{T}$.

Conditionnellement au fait que la cellule $\mathbf{i}$ contient x parasites, le nombre de parasites $(Z_{\mathbf{i}0}, Z_{\mathbf{i}1})$ de ses deux cellules filles est donné par

$$\sum_{k=1}^{x} (Z_k^{(0)}(\mathbf{i}), Z_k^{(1)}(\mathbf{i})),$$

avec $(Z_k^{(0)}(\mathbf{i}), Z_k^{(1)}(\mathbf{i}))_{\mathbf{i}\in\mathbb{T},k\geq 1}$ suite iid distribuée comme un couple de v.a. $(Z^{(0)}, Z^{(1)})$. On note

$$0 < m_0 := \mathbb{E}(Z^{(0)}), \qquad 0 < m_1 := \mathbb{E}(Z^{(1)}).$$

Les comportements asymptotiques (pour la guérison, le nombre de cellules contaminées et les proportions) dépendent des domaines suivants :

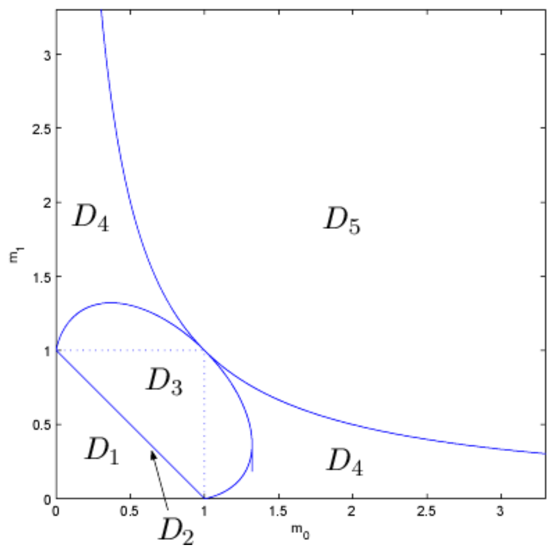
$D_1 = \{(m_0, m_1) : m_0 + m_1 < 1\}$

$D_2 = \{(m_0, m_1) : m_0 + m_1 = 1\}$

$D_3 = \{(m_0, m_1) : m_0 + m_1 > 1,$
$m_0 \log(m_0) + m_1 \log(m_1) < 0\}$

$D_4 = \{(m_0, m_1) : m_0 m_1 \leq 1,$
$m_0 \log(m_0) + m_1 \log(m_1) \geq 0\}$

$D_5 = \{(m_0, m_1) : m_0 m_1 > 1\}$



La guerison de l'organisme a lieu p.s. quand (m_0, m_1) se situe sous l'hyperbole.

Théorème. *Si $m_0 m_1 \leq 1$, alors $\#\mathbb{G}_n^*/\#\mathbb{G}_n \to 0$ quand $n \to \infty$.*
Sinon, $\#\mathbb{G}_n^*/\#\mathbb{G}_n \to 0$ ssi les parasites s'éteignent, ce qui se produit avec une probabilité inférieure à 1.

L'équivalent asymptotique du nombre de cellules infectées dépend des différents domaines. Nous séparons les résultats en deux parties, suivant que les parasites s'éteignent p.s. ou non.

Théorème. *Si (m_0, m_1) appartient à D_1, alors conditionnellement à la survie des parasites au temps n, $\#\mathbb{G}_n^*$ converge en distribution vers une variable aléatoire finie positive.*
Si (m_0, m_1) appartient à D_2, alors conditionnellement à la survie des parasites au temps n, $\#\mathbb{G}_n^*/n$ converge en distribution vers une variable aléatoire exponentielle.

De plus, en notant que le nombre total de parasites suit un processus de Galton Watson de moyenne $m_0 + m_1$:

Théorème. *Conditionnellement à la survie des parasites,*

Si (m_0, m_1) *appartient à* D_3*, alors* $\#\mathbb{G}_n^*/(m_0+m_1)^n$ *converge en probabilité vers une variable aléatoire finie positive quand* $n \to \infty$*.*

Si (m_0, m_1) *appartient à la frontière entre* D_3 *et* D_4*, alors* $\mathbb{E}[\#\mathbb{G}_n^*]/(n^{-1/2}(m_0+m_1)^n)$ *tends vers un nombre positif quand* $n \to \infty$*.*

Si (m_0, m_1) *appartient à* D_4*, alors* $\mathbb{E}[\#\mathbb{G}_n^*]/(n^{-3/2}\gamma^n)$ *converge vers un nombre positif quand* $n \to \infty$*, avec* $0 < \gamma < m_0 + m_1$*.*

Si (m_0, m_1) *appartient à la frontière de* D_5*, alors* $\mathbb{E}[\mathbb{G}_n^*]/(2^n n^{-1/2})$ *converge vers un nombre positif.*

Si (m_0, m_1) *appartient à l'intérieur de* D_5*, alors* $\mathbb{G}_n^*/2^n$ *converge p.s. vers une variable aléatoire finie positive.*

Nous pouvons ensuite donner le comportement asymptotique des proportions de cellules infectées par un nombre donné de parasites,

$$F_k(n) = \frac{\#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} = k\}}{\#\mathbb{G}_n^*}, \qquad (n, k \geq 1).$$

Théorème. *Si* (m_0, m_1) *appartient à* D_3*, alors pour tout* $k \geq 1$*,* $F_k(n) \to \mathbb{P}(\Upsilon = k)$ *en probabilité quand* $n \to \infty$*, où* Υ *est la limite quasi-stationaire du nombre de parasites dans une lignée cellulaire aléatoire.*

Si (m_0, m_1) *appartient à* D_3*, alors pour tout* $k \geq 1$*,* $F_k(n) \to 0$ *en probabilité quand* $n \to \infty$*.*

<u>Contamination de cellules en division</u> Nous ajoutons au modèle précédent une contamination des cellules par un nombre aléatoire de parasites, qui dépend (uniquement) du fait que la cellule est déjà infectée ou non. On note Y_0 (resp. Y_1) la v.a. donnant la loi du nombre de parasites qui contamine une cellule non infectée (resp infectée) au cours d'une génération.

Nous considérons ici le comportement asymptotique des proportions de cellules contenant un nombre donné de parasites,

$$F_k(n) = \frac{\#\{\mathbf{i} \in \mathbb{G}_n : Z_{\mathbf{i}} = k\}}{2^n}, \qquad (n, k \geq 1).$$

Théorème. *Si* $m_0 m_1 < 1$ *et* $\max(\mathbb{E}(\log^+(Y_i)) : i = 0, 1) < \infty$*, alors pour tout* $k \in \mathbb{N}$*,* $F_k(n) \to f_k$ *quand* $n \to \infty$*, avec* $f_k \geq 0$ *tels que* $\sum_{k=0}^{\infty} f_k = 1$*.*
Sinon, pour tout $k \in \mathbb{N}$*,* $F_k(n) \to 0$ *en probabilité quand* $n \to \infty$*.*

En fait, nous démontrons ce résultat pour un modèle plus général de multiplication-répartition des parasites, où ce mécanisme est tiré de manière iid pour chaque cellule (voir Section 9.1). Voici un exemple.

Exemple : Répartition binomiale aléatoire des parasites. Nous considérons ici le cas où les parasites se reproduisent en suivant un processus de Galton Watson de loi de reproduction Z. La répartition des parasites est binomiale de paramètre $P \in [0,1]$ p.s. C'est-à-dire que pour chaque cellule on tire un réel $p \in [0,1]$ suivant la loi P, puis chaque parasite est envoyé dans la première cellule fille avec probabilité p, et dans la seconde cellule fille sinon. En l'absence de contamination, le critère de guérison p.s. de l'organisme est

$$\log(\mathbb{E}(Z)) \leq \mathbb{E}(\log(1/P)).$$

Dans le cas où il y a contamination par des parasites extérieures à la population de cellules, les proportions asympotiques de cellules avec un nombre donné de parasites sont non dégénérées (i.e. de somme égale à 1) ssi

$$\log(\mathbb{E}(Z)) < \mathbb{E}(\log(1/P)), \qquad \max(\mathbb{E}(\log^+(Y_i)) : i = 0, 1) < \infty.$$

Sinon, toutes les proportions asymptotiques sont nulles.

Part I

Recouvrement aléatoire pour le stockage de données

Chapter 2

Preliminaries

In this chapter, we give some results about the main two notions we will use to study the model of data storage introduced in the next section. First, we focus on regenerative sets on the half line, then on Lévy processes.

Throughout the first part, we use the classical notation δ_x for the Dirac mass at x and $\mathbb{N} = \{1, 2, \ldots\}$. If $\mathcal{R}$ is a measurable subset of $\mathbb{R}$, we denote by $\mid \mathcal{R} \mid$ its Lebesgue measure and by $\mathcal{R}^{cl}$ its closure. For every $x \in \mathbb{R}$, we denote by $\mathcal{R} - x$ the set $\{y - x : y \in \mathcal{R}\}$ and

$$g_x(\mathcal{R}) = \sup\{y \leq x : y \in \mathcal{R}\}, \qquad d_x(\mathcal{R}) = \inf\{y > x : y \in \mathcal{R}\}. \tag{2.1}$$

By convention, $\sup \varnothing = -\infty$ and $\inf \varnothing = +\infty$.

If I is a closed interval of $\mathbb{R}$, we denote by $\mathcal{H}(I)$ the space of closed subsets of I. For every $x \in \mathbb{R}$ and $A \subset \mathbb{R}$ we define

$$d(x, A) = \inf\{1 - e^{-|x-y|} : y \in A\},$$

and we endow $\mathcal{H}(I)$ with the Hausdorff distance d_H defined for all $A, B \in \mathcal{H}(I)$ by:

$$d_H(A, B) = \max\big(\sup_{x \in A} d(x, B), \sup_{x \in B} d(x, A)\big).$$

The topology induced by this distance is the topology of Matheron [72] : a sequence $\mathcal{R}_n$ in $\mathcal{H}(I)$ converges to $\mathcal{R}$ iff for each open set G and each compact set K,

$$\begin{aligned} &\mathcal{R} \cap G \neq \varnothing \quad \text{implies} \quad \mathcal{R}_n \cap G \neq \varnothing \quad \text{for } n \text{ large enough}, \\ &\mathcal{R} \cap K = \varnothing \quad \text{implies} \quad \mathcal{R}_n \cap K = \varnothing \quad \text{for } n \text{ large enough}. \end{aligned}$$

It is also the topology induced by the Hausdorff metric on a compact set using $\arctan(\mathcal{R} \cup \{-\infty, \infty\})$ or the Skorokhod metric using the class of 'descending saw-tooth functions' (see [72] and [40] for details).

2.1 Regenerative sets on the half line

In this section, we define regenerative sets, state their strong Markov property and a representation theorem.

Let Ω^0 be the space of closed subset of $[0,\infty]$ containing 0, which we endow with

$$\mathcal{G}^0 := \sigma(d_s, s \geq 0).$$

Let $\mathcal{R}$ be a random closed set of $[0,\infty)$ containing 0 a.s :

$$\mathcal{R} : (\Omega, \mathcal{G}) \longrightarrow (\Omega^0, \mathcal{G}^0) \quad \text{is measurable,}$$

with $\mathcal{G} = \sigma(d_s(\mathcal{R}), s \geq 0)$.

Following Maisonneuve [71], we call a random closed set $\mathcal{R} \subset [0,\infty)$ regenerative if it contains 0 and for every $x \geq 0$, conditionally on $\{d_x(\mathcal{R}) < \infty\}$, the random set $(\mathcal{R} - d_x(\mathcal{R})) \cap [0,\infty)$ is distributed as $\mathcal{R}$ and is independent of $[0, d_x(\mathcal{R})] \cap \mathcal{R}$, i.e.

$$\forall x \geq 0, \quad \mathbb{P}((\mathcal{R} - D(\mathcal{R}, x)) \cap [0,\infty) \in . \mid [0, d_x(\mathcal{R})] \cap \mathcal{R}) = \mathbb{P}(\mathcal{R} \in .) \tag{2.2}$$

on $\{d_x(\mathcal{R}) < \infty\}$.

We introduce the following filtration

$$\mathcal{G}_x := \sigma(d_t(\mathcal{R}), t \leq x).$$

In this section, we always work on $[0,\infty)$ and, for any set $R \subset [0,\infty)$, we will write R instead of $R \cap (0,\infty)$. We immediately have a strong regeneration property [71] :

Proposition 2.1.1. *For every $(\mathcal{G}_x^+)_{x \geq 0}$ stopping time T, we have :*

$$\mathbb{P}(\mathcal{R} - d_T(\mathcal{R}) \in . \mid \mathcal{G}_T^+) = \mathbb{P}(\mathcal{R} \in .) \tag{2.3}$$

on $\{d_T(\mathcal{R}) < \infty\}$.

proof. If T is a simple stopping time, (2.2) implies that for every measurable function f :

$$\mathbb{E}(f(\mathcal{R} - d_T(\mathcal{R})))|\mathcal{G}_T) = \mathbb{E}(f(\mathcal{R}))$$

We can now prove (2.3) for $f = g(d_{x_1}, d_{x_2}, .., d_{x_k})$ with g continuous.
Let $T_n \downarrow T$ with $T_n > T$ on $\{T < \infty\}$ and T_n simple stopping time, then :

$$\mathbb{E}(f(\mathcal{R} - d_{T_n}(\mathcal{R})) \mid \mathcal{G}_{T_n}) \longrightarrow \mathbb{E}(f(\mathcal{R} - d_T(\mathcal{R})) \mid \mathcal{G}_T^+)$$

since $x \mapsto d_{x_1}(\mathcal{R} - d_x(\mathcal{R})) = d_{d_x(\mathcal{R})+x_1}(\mathcal{R})$ is cad and g continuous.
The result follows by an argument of monotone class. $\square$

Here is the fundamental result [71]. We give here an elementary new proof, which is in the same vein as [44].

Theorem 2.1.2. *The closed range $\mathcal{R}$ of a subordinator $(S_x)_{x\geq 0}$ is a regenerative random subset of $[0,\infty)$. Moreover, every regenerative random subset of $[0,\infty)$ has the same distribution as the closed range of some subordinator, whose Laplace exponent (or Lévy data) is uniquely determined up to a multiplicative constant.*

In the proof, we will note νA instead of $\nu(A)$ when ν is a measure on $\mathbb{R}^+$ and A is a subset of $\mathbb{R}^+$.

proof. First we prove that the closed range $\mathcal{R}$ of a subordinator $(S_x)_{x\geq 0}$ is regenerative. Let

$$L_x = \inf\{u : S_u > x\} \qquad (x \geq 0).$$

As $d_x(\mathcal{R}) = S_{L_x}$, $\mathcal{R} - d_x(\mathcal{R})$ is the closed range of the subordinator $(S_{L_x+u} - S_{L_x})_{u\geq 0}$ $(\infty - \infty = \infty)$ and the result follows from independence and stationarity of increments of a subordinator.

Second, we prove that every regenerative set $\mathcal{R}$ is the closed range of some subordinator. Let us construct this subordinator. It is easy if its Lebesgue measure is not zero (see forthcoming case 1). Define

$$S_x := \inf\{u : \ |\mathcal{R} \cap [0,u]| > x\}, \quad (x \geq 0).$$

Note that $(S_x)_{x\geq 0}$ is $(\mathcal{G}_{S_x})_{x\geq 0}$ measurable and for every $x \geq 0$, S_x $(\mathcal{G}^+_x)_{x\geq 0}$ stopping time. Moreover, $(S_x)_{x\geq 0}$ is cadlag and nondecreasing.
For every $x \geq 0$, if $S_x < \infty$ then $d_{S_x}(\mathcal{R}) = S_x$, so $S_{x+y} - S_x = \inf\{u : |(\mathcal{R} - d_{S_x}(\mathcal{R}))| \cap [0,u] > y\}$. Using the strong regeneration property in S_x, we get that $S_{x+y} - S_x$ has the same distribution as S_y and is independent of $\mathcal{G}^+_{S_x}$.

By independence and stationarity of increments, either $S_0 = 0$ a.s (case 1) or $S_0 = \infty$ a.s (case 2).

CASE 1 : We assume $S_0 = 0$ a.s. Then $(S_x)_{x\geq 0}$ is a subordinator. We shall prove that its range is equal to $\mathcal{R}^- := \{x : d_x(\mathcal{R}) = x\}$

The range of the subordinator is included in $\mathcal{R}^-$, since $d_{S_x}(\mathcal{R}) = S_x$.
Using again the strong regeneration property in T stopping time taking values in $\mathcal{R}^-$, we get that $\inf\{u > T : \ |\mathcal{R} \cap [T,u]| > 0\} = 0$ a.s, then $T = S_{|\mathcal{R}\cap[0,T]|}$, which

gives the other inclusion.

CASE 2 : We assume here $S_0 = \infty$ a.s. Then $|\mathcal{R}| = 0$ and we search a subordinator whose drift is equal to 0. It remains to characterize the Lévy measure ν of this subordinator.

In that purpose, we consider the successive jumps of $\mathcal{R}$ of size strictly larger than $x > 0$. Let

$\tau_1^x := \inf\{u \,:\, d_u(\mathcal{R}) - u > x\}$ and $\tau_{i+1}^x := \inf\{u \geq d_{\tau_i^t}(\mathcal{R}) \,:\, d_u(\mathcal{R}) - u > x\}$ $(i \geq 1)$.

Note that for every $x \geq 0$, $(\tau_i^x)_{i\geq 1}$ are $(\mathcal{G}_u^0)_{u\geq 0}$ stopping times and define

$$r_i^x := d_{\tau_i^x}(\mathcal{R}) - \tau_i^x \qquad (\infty - \infty = \infty).$$

As $|\mathcal{R}|$, we have the following identities :

$$\tau_1^x = \lim_{\epsilon \to 0} \sum_{k < \inf\{i \,:\, r_i^\epsilon > x\}} r_k^\epsilon \tag{2.4}$$

$$d_x(\mathcal{R}) = \lim_{\epsilon \to 0} \sum_{i \in \mathbb{N}} r_i^\epsilon \; 1_{\sum_{k=0}^{i-1} r_k^\epsilon \leq x} \tag{2.5}$$

Moreover, by the strong regeneration property in τ_i^x, $(r_i^x : i \in \mathbb{N})$ are iid.

Let x_0 such that $\mathbb{P}(\tau_1^{x_0} < \infty) > 0$, then for every $x < x_0$, $\mathbb{P}(r_1^x \geq x_0) > 0$ and we introduce the measure ν on $[0, \infty)$ defined by $\nu(x_0, \infty) = 1$ and :

$$\nu(x, \infty) = \frac{1}{\mathbb{P}(r_1^x > x_0)} \text{ if } 0 < x < x_0 \quad \text{and} \quad \nu(x, \infty) = \mathbb{P}(r_1^{x_0} > x) \text{ if } x > x_0. \tag{2.6}$$

For all $x_1 \leq x_2 < x_3$ then

$$\begin{aligned}
\mathbb{P}(r_1^{x_2} > x_3) &= \sum_{k\geq 1} \mathbb{P}(r_1^{x_1} \leq x_2, .., r_{k-1}^{x_1} \leq x_2, r_k^{x_1} > x_3) \\
&= \sum_{k\geq 1} \mathbb{P}(r_1^{x_1} \leq x_2)^{k-1} \mathbb{P}(r_1^{x_1} > x_3) \\
&= \frac{\mathbb{P}(r_1^{x_1} > x_3)}{1 - \mathbb{P}(r_1^{x_1} \leq x_2)} \\
&= \frac{\mathbb{P}(r_1^{x_1} > x_3)}{\mathbb{P}(r_1^{x_1} > x_2)}.
\end{aligned}$$

Applying this identity to $x_0 \leq x < \widetilde{x}$, $x \leq \widetilde{x} < x_0$ and $x \leq x_0 < \widetilde{x}$ ensures that for all $x, \widetilde{x} \geq 0$,

$$\mathbb{P}(r_1^x > \widetilde{x}) = \frac{\nu(\widetilde{x}, \infty)}{\nu(x, \infty)}.$$

To introduce a subordinator of Lévy measure ν, we must check an integrability assumption. In that purpose, let $(T_n, \widetilde{r}_n)$ be a PPP with intensity measure $|.| \times \nu$. We call $\widetilde{r}_n^x$ its successive jumps strictly larger than $x > 0$ and define $\widetilde{\tau}_1^x$ by (2.4). Then $(\widetilde{r}_n^x : n \in \mathbb{N})$ are iid and and distributed as $(r_n^x : n \in \mathbb{N})$.
So $\widetilde{\tau}_1^x$ is distributed as τ_1^x and $\mathbb{E}\big[\exp(-\widetilde{\tau}_1^{x_0})\big] > 0$.
Let $T = \inf\{T_k : \widetilde{r}_k > x_0\}$ be the instant of the first jump of the PPP strictly larger than x_0, then

$$\begin{aligned}
\mathbb{E}\big[\exp(-\widetilde{\tau}_1^{x_0})\big] &= \mathbb{E}\Big[\exp(-\sum_{n\in\mathbb{N}} \widetilde{r}_n 1_{\{T_n<T\}})\Big] \\
&= \int \mathbb{P}(T \in dt)\mathbb{E}\Big[\exp(-\sum_{n\in\mathbb{N}} \widetilde{r}_n 1_{\{T_n<T\}})|T = t\Big] \\
&= \int_0^\infty \nu(x_0,\infty)e^{-t\nu(x_0,\infty)}e^{-t\int_0^{x_0}(1-e^{-y})\nu(dy)}dt
\end{aligned}$$

So $\displaystyle\int_0^\infty (1 - e^{-y})\nu(dy) < \infty$ which implies

$$\int_0^\infty \min(1, z)\nu(dz) < \infty.$$

We can then introduce a subordinator of Lévy data $(0,\nu)$, denoted by $(\widetilde{S}_x)_{x\geq 0}$, and its closed range $\widetilde{\mathcal{R}}$. As $|\widetilde{\mathcal{R}}|$=0 and for every $x \geq 0$, $(\widetilde{r}_n^x : n \in \mathbb{N})$ is distributed as $(r_n^x : n \in \mathbb{N})$, then by 2.5, $(d_x(\mathcal{R}) : x \geq 0)$ is distributed as $(d_x(\widetilde{\mathcal{R}}) : x \geq 0)$. Thus $\widetilde{\mathcal{R}}$ is distributed as $\mathcal{R}$. This ensures that $\mathcal{R}$ is distributed as the closed range of a subordinator of Lévy data $(0,\nu)$.

Finally we prove the uniqueness. Let $\mathcal{R}$ and $\widetilde{\mathcal{R}}$ be the closed ranges of two subordinators $(S_x)_{x\geq 0}$ and $(\widetilde{S}_x)_{x\geq 0}$ whose Lévy data are resp. (d,ν) and $(\widetilde{\mathrm{d}}, \widetilde{\nu})$. Assume that $\mathcal{R}$ and $\widetilde{\mathcal{R}}$ are identically distributed and let us prove that (d,ν) and $(\widetilde{\mathrm{d}}, \widetilde{\nu})$ are equal to constant multiples. Note that for every x, $|\mathcal{R}\cap[0, x]| = \mathrm{d}\widetilde{L}x$ and $|\widetilde{\mathcal{R}} \cap [0, x]| = \widetilde{\mathrm{d}}L_x$ (see Proposition 1.8 in [20]) have the same distribution. Then,
Either $\mathrm{d}=\widetilde{\mathrm{d}} = 0$, as for every $x > 0$, $(\widetilde{r}_n^x : n \in \mathbb{N})$ is distributed as $(r_n^x : n \in \mathbb{N})$ by definition, then ν and $\widetilde{\nu}$ are equal to a positive factor (use 2.6).
Or $\mathrm{d}= \widetilde{\mathrm{d}} > 0$, using Notations of case 2, (2.6) ensures again that ν and $\widetilde{\nu}$ are equal to a positive factor. Moreover, thanks to [20] :

$$\mathbb{P}(|\mathcal{R} \cap [0, \tau_1^{x_0}]| \leq 1) = \mathbb{P}(\mathrm{d}T \leq 1) = 1 - e^{-\frac{\nu(x_0,\infty)}{\mathrm{d}}}$$

and using that this term just depends on the law of $\mathcal{R}$, we can conclude. □

We say that a regenerative set is *self-similar* if it has the self-similarity property:

$$\forall c > 0 \;\; c\mathcal{R} \stackrel{d}{=} \mathcal{R}$$

We know that $\mathcal{R}$ is the closed range of a subordinator of Lévy data (d,ν).
If $\mathcal{R}$ has some points in $]0, \infty[$, the self-similarity property implies that 0 is not isolated on the right and $\nu(0, \infty) = \infty$. Moreover this property implies that either $|\mathcal{R}| = \infty$ and $\mathcal{R} = \mathbb{R}^+$ a.s, or $|\mathcal{R}|$=0 a.s and d =0 (see [20]). In the last case, we can also describe ν and we have :

Theorem 2.1.3. *A regenerative set is self-similar iff it is the closed range of some stable subordinator or the trivial set $\mathbb{R}^+$.*

proof. $\Rightarrow$ We use the discussion above and consider the case d=0.
The self-similarity property implies that for every $x \geq 0$, $\mathbb{P}(r_1^x > \widetilde{x}) = \mathbb{P}(r_1^{xc} > \widetilde{x}c)$ so, for $c > 0$ we have :

$$\frac{\nu(cx, \infty)}{\nu(x, \infty)} = \text{cst} = \nu(cx_0, \infty)$$

Putting $h(x) = \log(\nu(x_0 e^x, \infty))$ for $x \in \mathbb{R}$, we have $h(x + x') = h(x) + h(x')$. Moreover $\nu(x, \infty)$ is cad, so:

$$\nu(dx) = \frac{c}{x^{1+\alpha}} \mathrm{d}x$$

for some $\alpha \in \mathbb{R}$. Using that ν verifies

$$\int_0^\infty \min(1, z)\nu(\mathrm{d}z) < \infty,$$

and $\nu(0, \infty) = \infty$, we get that $0 < \alpha < 1$.

$\Leftarrow$ If $(S_x)_{x\geq 0}$ is a stable subordinator, then for every $r > 0$, $(S_{rx})_{x\geq 0} \stackrel{d}{=} (c/r^{1+\alpha} S_x)_{x\geq 0}$ which implies that $\forall c > 0,\;\; c\mathcal{R} \stackrel{d}{=} \mathcal{R}$. □

2.2 Background on Lévy processes

The results given in this section can be found in the Chapters VI and VII in [19] (there, statements are made in terms of the dual process $-Y$). We recall that a Lévy process is càdlàg process starting from 0 which has iid increments. A subordinator is an increasing Lévy process.

We consider a Lévy process $(X_x)_{x\geq 0}$ which has no negative jumps (spectrally positive Lévy process). We denote by Ψ its Laplace exponent which verifies for every $\rho \geq 0$:

$$\mathbb{E}(\exp(-\rho X_x)) = \exp(-x\Psi(\rho)). \tag{2.7}$$

We stress that this is not the classical choice for the sign of the Laplace exponent of Lévy processes with no negative jumps and a negative drift. However it is the classical choice for subordinators, which we will need. It is then convenient to use this same definition for all Lévy processes which appear in this text.

First, we consider the case when $(X_x)_{x\geq 0}$ has bounded variation. That is,

$$X_x := dx + \sum_{x_i \leq x} l_i,$$

where $\{(x_i, l_i) : i \in \mathbb{N}\}$ is a PPP on $[0,\infty[\times[0,\infty[$ with intensity measure $\mathrm{d}x \otimes \nu$ such that $\int_0^\infty x\nu(\mathrm{d}x) < \infty$. We call ν the Lévy measure and $d \in \mathbb{R}$ the drift. Note that $(X_x)_{x\geq 0}$ is a subordinator iff $d \geq 0$.

Writing $\bar{\nu}$ for the tail of the measure ν, the Lévy-Khintchine formula gives

$$\Psi(\rho) = d\rho + \int_0^\infty (1 - e^{-\rho x})\nu(\mathrm{d}x), \tag{2.8}$$

$$\frac{\Psi(\rho)}{\rho} = d + \int_0^\infty e^{-\rho x}\bar{\nu}(x)\mathrm{d}x, \tag{2.9}$$

$$\Psi'(0) = d + \int_0^\infty x\nu(\mathrm{d}x), \tag{2.10}$$

$$\lim_{\rho\to\infty} \frac{\Psi(\rho)}{\rho} = d \quad \text{and} \quad \lim_{\rho\to\infty} (\Psi(\rho) - d\rho) = \bar{\nu}(0). \tag{2.11}$$

Second, we consider the case when Ψ has a right derivative at 0 with

$$\Psi'(0) < 0, \tag{2.12}$$

meaning that $\mathbb{E}(X_1) < 0$. And we consider the infimum process which has continuous path and the first passage time defined for $x \geq 0$ by

$$I_x = \inf\{X_y : 0 \leq y \leq x\} \quad ; \qquad \tau_x = \inf\{z \geq 0 : X_z < -x\}.$$

As $-\Psi$ is strictly convex and $-\Psi'(0) > 0$, $-\Psi$ is strictly increasing from $[0,\infty[$ to $[0,\infty[$ and so is strictly positive on $]0,\infty]$. We write $\kappa : [0,\infty[\to \mathbb{R}$ for the inverse function of $-\Psi$ and we have (see [19] Theorem 1 on page 189 and Corollary 3 on page 190) :

Theorem 2.2.1. *$(\tau_x)_{x\geq 0}$ is a subordinator with Laplace exponent κ.*

Moreover the following identity holds between measures on $[0,\infty[\times[0,\infty[$:

$$x\mathbb{P}(\tau_l \in \mathrm{d}x)\mathrm{d}l = l\mathbb{P}(-X_x \in \mathrm{d}l)\mathrm{d}x. \tag{2.13}$$

Note that if $(X_x)_{x\geq 0}$ has bounded variations, using (2.11), we can write

$$\forall\ \rho \geq 0, \quad \kappa(\rho) = -\frac{\rho}{d} + \int_0^\infty (1 - e^{-\rho z})\Pi(\mathrm{d}z), \tag{2.14}$$

where Π is a measure on $\mathbb{R}^+$ verifying (use (2.11) and Wald's identity or (2.10)) :

$$\bar{\Pi}(0) = -\frac{\bar{\nu}(0)}{d}, \qquad \int_0^\infty x\Pi(\mathrm{d}x) = \frac{1}{d} - \frac{1}{d + \int_0^\infty x\nu(\mathrm{d}x)}. \tag{2.15}$$

Now we introduce the supremum process defined for $x \geq 0$ by

$$S_x := \sup\{X_y : 0 \leq y \leq x\},$$

and the a.s. unique instant at which X reaches this supremum on $[0,x]$:

$$\gamma_x := \inf\{y \in [0,x] : X_y = S_x\}.$$

By duality, we have $(S_x, \gamma_x) \stackrel{d}{=} (X_x - I_x, x - g_x)$ where g_x denotes the a.s unique instant at which $(X_{x-})_{x\geq 0}$ reaches its overall infimum on $[0,x]$ (see Proposition 3 in [19] or [21] on page 25). If T is an exponentially distributed random time with parameter $q > 0$ which is independent of X and $\lambda, \mu > 0$, then we have (use [19] Theorem 5 on page 160 and Theorem 4 on page 191) :

$$\begin{aligned}
\mathbb{E}\big(\exp(-\mu S_T - \lambda\gamma_T)\big) &= \frac{q(\kappa(\lambda+q) - \mu)}{\kappa(q)(q + \lambda + \Psi(\mu))} \\
&= \exp\Big(\int_0^\infty \mathrm{d}x \int_0^\infty \mathbb{P}(Y_x \in \mathrm{d}y)(e^{-\lambda x - \mu y} - 1)x^{-1}e^{-qx}\Big),
\end{aligned}$$

which gives

$$\mathbb{E}\big(\exp(-\mu S_\infty - \lambda\gamma_\infty)\big) = \frac{1}{\kappa'(0)}\frac{\kappa(\lambda) - \mu}{\lambda + \Psi(\mu)} = -\Psi'(0)\frac{\kappa(\lambda) - \mu}{\lambda + \Psi(\mu)}, \tag{2.16}$$

$$\mathbb{E}\big(\exp(-\mu S_\infty)\big) = \mu\frac{\Psi'(0)}{\Psi(\mu)}, \tag{2.17}$$

$$\mathbb{E}\big(\exp(-\lambda\gamma_\infty)\big) = \exp\Big(\int_0^\infty (e^{-\lambda x} - 1)x^{-1}\mathbb{P}(X_x > 0)\mathrm{d}x\Big). \tag{2.18}$$

Chapter 3

Poissonian model for data storage in continuous time

3.1 Introduction

We consider a generalized version in continuous time of the original parking problem of Knuth. Knuth was interested in the storage of data on a hardware represented by a circle with n spots. Files arrive successively at locations chosen uniformly at random and independently among these n spots. They are stored in the first free spot at the right of their arrival point (at their arrival point if it is free). Initially Knuth worked on the hashing of data (see e.g. [29, 37, 41]) : he studied the distance between the spots where the files arrive and the spots where they are stored. Later Chassaing and Louchard [28] have described the evolution of the largest block of data in such coverings when n tends to infinity. They observed a phase transition at the stage where the hardware is almost full, which is related to the additive coalescent. Bertoin and Miermont [23] have extended these results to files of random sizes which arrive uniformly on the circle.

We consider here a continuous time version of this model where the hardware is large and now identified with the real line. A file labeled i of length (or size) l_i arrives at time $t_i \geq 0$ at location $x_i \in \mathbb{R}$. The storage of this file uses the free portion of size l_i of the real line at the right of x_i as close to x_i as possible (see Figure 3). That is, it covers $[x_i, x_i + l_i[$ if this interval is free at time t_i. Otherwise this file can be splited into several parts which are then stored in the closest free portions at the right of the arrival location. We require uniformity of the location where they arrive and identical distribution of the sizes and we model the arrival of files by a Poisson point process (PPP) : $\{(t_i, x_i, l_i) \ : \ i \in \mathbb{N}\}$ is a PPP with intensity $\mathrm{d}t \otimes \mathrm{d}x \otimes \nu(\mathrm{d}l)$ on $\mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+$. We denote $\mathrm{m} := \int_0^\infty l\nu(\mathrm{d}l)$ and assume $\mathrm{m} < \infty$. So m is the mean of the sum of sizes of files which arrive during a unit interval time on some interval with unit length.

We begin by constructing this random covering (Section 3.2). The first questions which arise and are treated here concern statistics at a fixed time for the set of occupied locations. What is the distribution of the covering at a fixed time ? At what time the hardware becomes full ? What are the asymptotics of the covering at this saturation time ? What is the length of the largest block on a part of the hardware ?

It is quite easy to see that the hardware becomes full at a deterministic time equal to 1/m. In Section 3.3.1, we give some geometric properties of the covering and characterize the distribution of the covering $\mathcal{C}(t)$ at a fixed time t by giving the joint distribution of the block of data straddling 0 and the free spaces on both sides of this block. Results are stated in Section 3.3 and proved in Section 3.4. These results will be useful for the problem of the dynamic of the covering considered in Chapter 5, where we investigate the evolution in time of a typical data block. Moreover, using this characterization, we determine the asymptotic regimes at the saturation time, which depend on the tail of ν, as in [23, 27, 28] . More precisely, we can give in the next Chapter the asymptotic of set of occupied locations $\mathcal{C}(t)$ when t tends to 1/m (Theorem 4.1.1) and the asymptotic of $\mathcal{C}(t)$ restricted to $[0, x]$ when x tends to infinity and t tends to 1/m (Theorem 4.2.1). We derive then the asymptotic of the largest block of the hardware restricted to $[0, x]$ when x tends to infinity and t tends to 1/m. As expected, we recover the phase transition observed by Chassaing and Louchard in [28].

It is easy to check that for each fixed time t, $\mathcal{C}(t)$ does not depend on the order of arrival of files before time t. Thus, if ν is finite, we can view the files which arrive before time t as customers : the size of the file l becomes the service time of the customer and the location x where the file arrives becomes the arrival time of the customer. We are then in the framework of the $M/G/1$ queue model in the stationary regime and the covering $\mathcal{C}(t)$ becomes the union of busy periods (see e.g. Chap 3 in [30] or [81]). Thus, several results for finite ν follow easily from known results on $M/G/1$. When ν is infinite, results are similar though busy cycle is not defined. Thus the approach is different and proving asymptotics on random sets requires results about Lévy processes and regenerative sets. Moreover, as far as we know, the longest busy period and more generally asymptotic regimes on $[0, x]$ when x tends to infinity and t tends to the saturation time (Section 4.2) have not been considered in queuing model.

3.2 Construction

First, we present a deterministic construction of the covering $\mathcal{C}$ associated with a given sequence of files labelled by $i \in \mathbb{N}$. The file labelled by $i \in \mathbb{N}$ has size l_i and arrives after the files labelled by $j \leq i-1$, at location x_i on the real line. Files are stored following the process described in the Introduction and $\mathcal{C}$ is the portion of line which is used for the storage. We begin by constructing the covering $\mathcal{C}^{(n)}$ obtained by considering only the first n files, so that $\mathcal{C}$ is obtained as the increasing union of these coverings. A short thought (see Remark 1 p 5) enables us to see that the covering $\mathcal{C}$ does not depend on the order of arrival of the files. This construction of $\mathcal{C}$ will then be applied to the construction of our random covering at a fixed time $\mathcal{C}(t)$ by considering files arrived before time t. This construction and results for finite ν are classical in queuing theory (see e.g. [30]) and storage systems (see e.g. [81]). Thus we do not give details here and we refer to [11] for complete proof.

We define $\mathcal{C}^{(n)}$ by induction. We set $\mathcal{C}^{(0)} := \varnothing$, and introduce the complementary set $\mathcal{R}^{(n)}$ of $\mathcal{C}^{(n)}$ (i.e. the free space of the real line). Let $y_{n+1} = \inf\{y \geq 0, \mid \mathcal{R}^{(n)} \cap [x_{n+1}, y[| = l_{n+1}\}$, so y_{n+1} is the right-most point which is used for storing the $(n+1)$-th file. Define then

$$\mathcal{C}^{(n+1)} := \mathcal{C}^{(n)} \cup [x_{n+1}, y_{n+1}[.$$

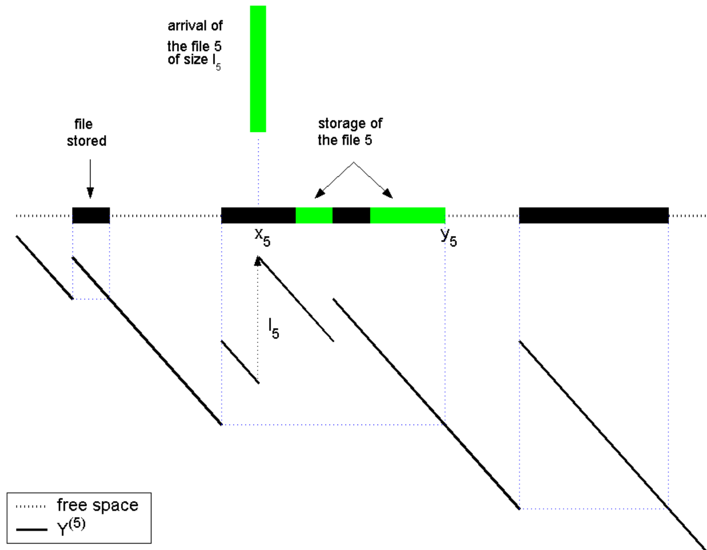
Now we consider the quantity of data over x, $R_x^{(n)}$, as the quantity of data which we have tried to store at the location x (successfully or not) when n files are stored. These data are the data fallen in $[g_x(\mathcal{R}^{(n)}), x]$ which could not be stored in $[g_x(\mathcal{R}^{(n)}), x]$, so $R_x^{(n)}$ is defined by

$$R_x^{(n)} := -(x - g_x(\mathcal{R}^{(n)})) \ + \sum_{\substack{i \leq n \\ x_i \in [g_x(\mathcal{R}^{(n)}), x]}} l_i.$$

Note that in queuing systems, $R^{(n)}$ is the workload. This quantity can be expressed using the function $Y^{(n)}$, which sums the sizes of the files arrived at the left of a point x minus the drift term x. It is defined by $Y_0^{(n)} = 0$ and

$$Y_b^{(n)} - Y_a^{(n)} = -(b-a) \ + \sum_{\substack{i \leq n \\ x_i \in]a,b]}} l_i \qquad \text{for } a < b. \tag{3.1}$$

Figure 3. Arrival and storage of the 5-file and representation of $Y^{(5)}$. The first four files have been stored without splitting and are represented by the black rectangles.



Introducing also its infimum function defined for $x \in \mathbb{R}$ by $I_x^{(n)} := \inf\{Y_y^{(n)} : y \leq x\}$, we get the following expression.

Lemma 3.2.1. *For every $n \geq 1$, we have $R^{(n)} = Y^{(n)} - I^{(n)}$.*

Proof. Let $x \in \mathbb{R}$. For every $y \leq x$, the quantity of data over x is at least the quantity of data fallen in $[y, x]$ minus $y - x$, i.e.

$$R_x^{(n)} \geq \sum_{\substack{i \leq n \\ x_i \in [y,x]}} l_i \; - (x - y)$$

and by definition of $R_x^{(n)}$, we get :

$$R_x^{(n)} = \sup\{ \sum_{\substack{i \leq n \\ x_i \in [y,x]}} l_i \; - \; (x-y) : \; y \leq x\} = \sup\{Y_x^{(n)} - Y_y^{(n)} : \; y \leq x\}.$$

Then $R_x^{(n)} = Y_x^{(n)} - I_x^{(n)}$. □

As a consequence, the covered set when the first n files are stored is given by

$$\mathcal{C}^{(n)} = \{x \in \mathbb{R} : \; Y^{(n)} - I^{(n)} > 0\}. \tag{3.2}$$

We are now able to investigate the situation when n tends to infinity under the following mild condition

$$\forall \; L \geq 0, \quad \sum_{x_i \in [-L,L]} l_i < \infty, \tag{3.3}$$

which means that the quantity of data arriving on a compact set is finite. Introduce the function Y defined on $\mathbb{R}$ by $Y_0 = 0$ and

$$Y_b - Y_a = -(b-a) \; + \sum_{x_i \in]a,b]} l_i \qquad \text{for } a < b,$$

and its infimum I defined for $x \in \mathbb{R}$ by $\quad I_x := \inf\{Y_y : y \leq x\}$.

As expected, letting $n \to \infty$ in (3.1) and (3.2), the covering

$$\mathcal{C} := \cup_{n \in \mathbb{N}} \mathcal{C}^{(n)}$$

is given by :

Proposition 3.2.2. - *If* $\lim_{x \to -\infty} Y_x = +\infty$, *then* $\mathcal{C} = \{x \in \mathbb{R} : \; Y_x - I_x > 0\} \neq \mathbb{R}$.

- *If* $\liminf_{x \to -\infty} Y_x = -\infty$, *then* $\mathcal{C} = \{x \in \mathbb{R} : \; Y_x - I_x > 0\} = \mathbb{R}$.

Remark 1. This result ensures that the covering at time t just depend on $(Y_x - I_x)_{x \in \mathbb{R}}$. Thus it does not depend on the order of arrival of files.

Proof. Condition (3.3) ensures that $Y^{(n)}$ converges to Y uniformly on every compact set of $\mathbb{R}$.

• If $\lim_{x \to -\infty} Y_x = +\infty$, then for every $L \geq 0$, there exists $L' \geq L$ such that $I_{-L'} = Y_{-L'}$. Moreover $Y_x \leq Y_x^{(n)}$ if $x \leq 0$. So :

$$Y_{-L'}^{(n)} \overset{n \to \infty}{\longrightarrow} Y_{-L'} = I_{-L'} \quad \text{and} \quad I_{-L'} \leq I_{-L'}^{(n)} \leq Y_{-L'}^{(n)}$$

Then $I_{-L'}^{(n)} \overset{n \to \infty}{\longrightarrow} I_{-L'}$. As $Y^{(n)}$ converges to Y uniformly on $[-L', L']$, this entails that for every x in $[-L, L]$, $\inf\{Y_y^{(n)}, -L' \leq y \leq x\} \overset{n \to \infty}{\longrightarrow} \inf\{Y_y, -L' \leq y \leq x\}$. Then,

$$I_x^{(n)} = I_{-L'}^{(n)} \wedge \inf\{Y_y^{(n)}, -L' \leq y \leq x\} \overset{n \to \infty}{\longrightarrow} I_{-L'} \wedge \inf\{Y_y, -L' \leq y \leq x\} = I_x.$$

So $Y_x^{(n)} - I_x^{(n)} \overset{n \to \infty}{\longrightarrow} Y_x - I_x$ and $Y_x^{(n)} - I_x^{(n)}$ increases when n increases since it is equal to $R_x^{(n)}$, the quantity of data over x (see Lemma 3.2.1). We conclude that there is the identity

$$\{x \in \mathbb{R}, \; Y_x - I_x > 0\} = \cup_{n \in \mathbb{N}} \{x \in \mathbb{R}, \; Y_x^{(n)} - I_x^{(n)} > 0\} = \mathcal{C}.$$

Moreover $-L' \notin \{x \in \mathbb{R} : \ Y_x - I_x > 0\}$, so $\mathcal{C} = \{x \in \mathbb{R}, \ Y_x - I_x > 0\} \neq \mathbb{R}$.

- If $\liminf_{x \to -\infty} Y_x = -\infty$, then for every $x \in \mathbb{R}$,

$$I_x = -\infty \qquad \text{and} \qquad I_x^{(n)} \overset{n\to\infty}{\longrightarrow} -\infty$$

The first identity entails that $\{x \in \mathbb{R}, \ Y_x - I_x > 0\} = \mathbb{R}$. As $(Y_x^{(n)})_{n\in\mathbb{N}}$ is bounded, the second one implies that there exists n in $\mathbb{N}$ such that $Y_x^{(n)} - I_x^{(n)} > 0$. Then we have also $\cup_{n\in\mathbb{N}}\{x \in \mathbb{R} : \ Y^{(n)} - I^{(n)} > 0\} = \mathbb{R}$, which gives the result. $\square$

Finally, we can construct the random covering associated with a PPP. As the order of arrival of files has no importance, the random covering $\mathcal{C}(t)$ at time t described in Introduction is obtained by the deterministic construction above by taking the subfamily of files i which verifies $t_i \leq t$.

When files arrive according to a PPP, $(Y_x)_{x\geq 0}$ is a Lévy process and we can use results of Section 2.2.

3.3 Properties at a fixed time.

3.3.1 Statistics at a fixed time

Our purpose in this section is to specify the distribution of the covering $\mathcal{C}(t)$ using the characterization of Section 3.2 and results of Section 2.2. This characterization will be useful to prove asymptotics results (Theorem 4.1.1, Theorem 4.2.1 and Corollary 4.2.2) and for the dynamic results given in Chapter 5. In that view, following the previous section, we consider the process $(Y_x^{(t)})_{x\in\mathbb{R}}$ associated to the PPP $\{(t_i, l_i, x_i), i \in \mathbb{N}\}$ defined by

$$Y_0^{(t)} := 0 \quad ; \quad Y_b^{(t)} - Y_a^{(t)} = -(b-a) \ + \sum_{\substack{t_i \leq t \\ x_i \in]a,b]}} l_i \quad \text{for } a < b,$$

which has independent and stationary increments, no negative jumps and bounded variation. Introducing also its infimum process defined for $x \in \mathbb{R}$ by

$$I_x^{(t)} := \inf\{Y_y^{(t)} : y \leq x\},$$

we can give now a handy expression for the covering at a fixed time and obtain that the hardware becomes full at a deterministic time equal to 1/m (see Section 3.4 for the proof).

Proposition 3.3.1. *For every $t < 1/\mathrm{m}$, we have $\mathcal{C}(t) = \{x \in \mathbb{R} : Y_x^{(t)} > I_x^{(t)}\} \neq \mathbb{R}$ a.s.*
For every $t \geq 1/\mathrm{m}$, we have $\mathcal{C}(t) = \mathbb{R}$ a.s.

Indeed, in queuing system, $t\mathrm{m}$ is the charge and $\mathcal{C}(t) \neq \mathbb{R} \;\Leftrightarrow\; t\mathrm{m} < 1$ is the standard claim of stability for $tm < 1$. The complete argument is deferred to Section 3.4.

To specify the distribution of $\mathcal{C}(t)$, it is equivalent and more convenient to describe its complementary set, denoted by $\mathcal{R}(t)$, which corresponds to the free space of the hardware. By the previous proposition, there is the identity :

$$\mathcal{R}(t) = \{x \in \mathbb{R} : Y_x^{(t)} = I_x^{(t)}\}. \tag{3.4}$$

We begin by giving some geometric properties of this set. These properties are classical (see [56] for storage systems and [62] for queuing theory).

Proposition 3.3.2. *For every $t \geq 0$, $\mathcal{R}(t)$ is stationary, its closure is symmetric in distribution and it enjoys the regeneration property :*
For every $x \in \mathbb{R}$, $(\mathcal{R}(t) - d_x(\mathcal{R}(t))) \cap [0, \infty[$ is independent of $\mathcal{R}(t) \cap]-\infty, x]$ and is distributed as $(\mathcal{R}(t) - d_0(\mathcal{R}(t))) \cap [0, \infty[$.
Moreover for every $x \in \mathbb{R}$, $\mathbb{P}(x \in \mathcal{C}(t)) = \min(1, \mathrm{m}t)$.

Stationarity is plain from the construction of the covering and regeneration property is a direct consequence of Lemma 3.4.1. Symmetry is then a consequence of Lemma 6.5 in [89] or Corollary (7.19) in [90]. Computation of $\mathbb{P}(x \in \mathcal{C}(t))$ is then derived from Theorem 1 in [89]. See the next section for details.

Even though for each fixed t the distribution of $\mathcal{R}(t)^{cl}$ is symmetric, the processes $(\mathcal{R}(t)^{cl} : t \in [0, 1/\mathrm{m}])$ and $(-\mathcal{R}(t)^{cl} : t \in [0, 1/\mathrm{m}])$ are quite different. For example, we shall observe in [12] that the left extremity of the data block straddling 0 is a Markov process but the right extremity is not.

We want now to characterize the distribution of the free space $\mathcal{R}(t)$. For this purpose, we need some notation. The drift of the Lévy process $(Y_x^{(t)})_{x \geq 0}$ is equal to -1, its Lévy measure is equal to $t\nu$ and its Laplace exponent $\Psi^{(t)}$ is then given by (see (2.8))

$$\Psi^{(t)}(\rho) := -\rho + \int_0^\infty \left(1 - e^{-\rho x}\right) t\nu(\mathrm{d}x). \tag{3.5}$$

For sake of simplicity, we write, recalling (2.1),

$$g(t) := g_0(\mathcal{R}(t)), \qquad d(t) = d_0(\mathcal{R}(t)), \qquad l(t) = d(t) - g(t),$$

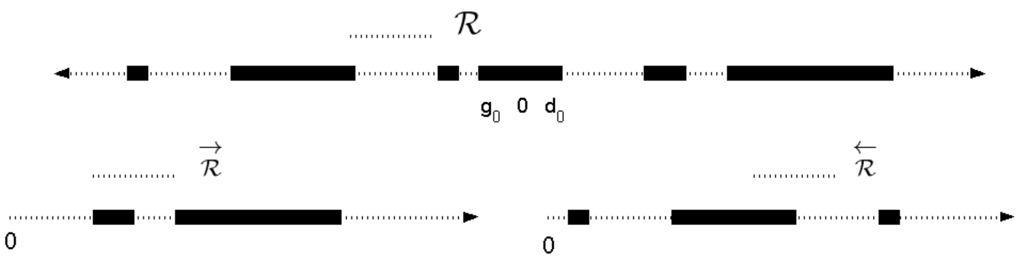
which are respectively the left extremity, the right extremity and the length of the data block straddling 0, $\mathbf{B}_0(t)$. Note that $g(t) = d(t) = 0$ if $\mathbf{B}_0(t) = \varnothing$.

We work with $\mathcal{R}$ subset of $\mathbb{R}$ of the form $\sqcup_{n\in\mathbb{N}}[a_n, b_n[$ and we denote by $\widetilde{\mathcal{R}} := \sqcup_{n\in\mathbb{N}}[-b_n, -a_n[$ the symmetrical of $\mathcal{R}$ with respect to 0 closed at the left, open at the right. We consider the positive part (resp. negative part) of $\mathcal{R}$ defined by

$$\overrightarrow{\mathcal{R}} := (\mathcal{R} - d_0(\mathcal{R})) \cap [0, \infty] = \bigsqcup_{n\in\mathbb{N}:\ a_n \geq d_0(\mathcal{R})} [a_n - d_0(\mathcal{R}), b_n - d_0(\mathcal{R})[,$$

$$\overleftarrow{\mathcal{R}} := \overrightarrow{\widetilde{\mathcal{R}}} = \bigsqcup_{n\in\mathbb{N}:\ b_n \leq g_0(\mathcal{R})} [g_0(\mathcal{R}) - b_n, g_0(\mathcal{R}) - a_n[.$$

Example 1. For a given $\mathcal{R}$ represented by the dotted lines, we give below $\overrightarrow{\mathcal{R}}$ and $\overleftarrow{\mathcal{R}}$, which are also represented by dotted lines. Moreover the endpoints of the data blocks containing 0 are denoted by g_0 and d_0.



Thus $\overrightarrow{\mathcal{R}(t)}$ (resp. $\overleftarrow{\mathcal{R}(t)}$) is the free space at the right of $\mathbf{B}_0(t)$ (resp. at the left of $\mathbf{B}_0(t)$, turned over, closed at the left and open at the right). We have then the identity

$$\mathcal{R}(t) = (d(t) + \overrightarrow{\mathcal{R}(t)}) \sqcup (\widetilde{-g(t) + \overleftarrow{\mathcal{R}(t)}}). \tag{3.6}$$

Introducing also the processes $(\overrightarrow{\tau}^{(t)}_x)_{x\geq 0}$ and $(\overleftarrow{\tau}^{(t)}_x)_{x\geq 0}$ defined by

$$\overrightarrow{\tau}^{(t)}_x := \inf\{y \geq 0 : |\overrightarrow{\mathcal{R}(t)} \cap [0, y]| > x\}, \qquad \overleftarrow{\tau}^{(t)}_x := \inf\{y \geq 0 : |\overleftarrow{\mathcal{R}(t)} \cap [0, y]| > x\},$$

enables us to describe $\mathcal{R}(t)$ in the following way (see Section 3.4 for the proof).

Proposition 3.3.3. *(i) The random sets $\overrightarrow{\mathcal{R}(t)}$ and $\overleftarrow{\mathcal{R}(t)}$ are independent, identically distributed and independent of $(g(t), d(t))$.*
(ii) $\overrightarrow{\mathcal{R}(t)}$ and $\overleftarrow{\mathcal{R}(t)}$ are the range of the subordinators $\overrightarrow{\tau}^{(t)}$ and $\overleftarrow{\tau}^{(t)}$ respectively whose Laplace exponent $\kappa^{(t)}$ is the inverse function of $-\Psi^{(t)}$.
(iii) The distribution of $(g(t), d(t))$ is specified by :

$$(g(t), d(t)) = (-Ul(t), (1 - U)l(t)),$$

$$\mathbb{P}(l(t) \in \mathrm{d}x) = (1 - \mathrm{m}t)\big(\delta_0(\mathrm{d}x) + \mathbb{1}_{\{x>0\}} x \Pi^{(t)}(\mathrm{d}x)\big),$$

where U is an uniform random variable on $[0, 1]$ independent of $l(t)$ and $\Pi^{(t)}$ is the Lévy measure of $\kappa^{(t)}$.

Remark 2. Such results are classical for regenerative sets (see e.g. [56, 71, 79]). But we need this particular characterization and expressions given in the proof below for forthcoming results.

We can then estimate the number of data blocks on the hardware. If ν has a finite mass, we write $N_x^{(t)}$ the number of data blocks of the hardware restricted to $[-x, x]$ at time t. This quantity has a deterministic asymptotic as x tends to infinity which is maximum at time $1/(2\mathrm{m})$. And the number of blocks of the hardware reaches a.s. its maximal at time $1/(2\mathrm{m})$. More precisely,

Corollary 3.3.4. *If $\bar{\nu}(0) < \infty$, then for every $t \in [0, 1/\mathrm{m}[$,*

$$\lim_{x \to \infty} \frac{N_x^{(t)}}{2x} = \bar{\nu}(0) t (1 - \mathrm{m}t) \quad a.s.$$

Moreover, we can describe here the hashing of data. We recall that a file labeled by i is stored at location x_i. In the hashing problem, one is interested by the location where the file i is stored knowing x_i. By stationarity, we can take $x_i = 0$ and consider a file of size l which we store at time t at location 0 on the hardware whose free space is equal to $\mathcal{R}(t)$. The first point (resp. the last point) of the hardware occupied for the storage of this file is equal to $d(t)$ (resp. to $d(t) + \vec{\tau}_l^{(t)}$). This gives the distribution of the extremities of the portion of the hardware used for the storage of a file.

3.3.2 Observations and examples

First, we have for every $\rho \geq 0$ (use (2.14)),

$$\kappa^{(t)}(\rho) = \rho + \int_0^\infty (1 - e^{-\rho x}) \Pi^{(t)}(\mathrm{d}x), \tag{3.7}$$

and using (2.15)

$$\bar{\Pi}^{(t)}(0) = t\bar{\nu}(0), \qquad \int_0^\infty x \Pi^{(t)}(\mathrm{d}x) = \frac{\mathrm{m}t}{1 - \mathrm{m}t}. \tag{3.8}$$

Using (2.13), we have also the following identity of measures on $[0, \infty[\times[0, \infty[$

$$x \mathbb{P}(\vec{\tau}_l^{(t)} \in \mathrm{d}x)\mathrm{d}l = l\mathbb{P}(-Y_x^{(t)} \in \mathrm{d}l)\mathrm{d}x. \tag{3.9}$$

Finally, we give the distribution of the extremities of $\mathbf{B}_0$:

$$\mathbb{P}(-g(t) \in \mathrm{d}x) = \mathbb{P}(d(t) \in \mathrm{d}x) = (1 - \mathrm{m}t)\big(\delta_0(\mathrm{d}x) + \mathbb{1}_{\{x>0\}} \bar{\Pi}^{(t)}(x)\mathrm{d}x\big). \tag{3.10}$$

Let us consider three explicit examples

Example 2. (1) The basic example is when $\nu = \delta_1$ (all files have the same unit size as in the original parking problem in [28]). Then for all $x \in \mathbb{R}_+$ and $n \in \mathbb{N}$,

$$\mathbb{P}(Y_x^{(t)} + x = n) = e^{-tx}\frac{(tx)^n}{n!},$$
$$\mathbb{P}(\overrightarrow{\tau}_x^{(t)} = x + n) = \frac{x}{x+n}e^{-t(x+n)}\frac{(t(n+x))^n}{n!}, \tag{3.11}$$

where the second identity follows from integrating (3.9) on $\{(x,l) : l \in [z, z+h], x - z = n\}$ and letting h tend to 0. Then,

$$\Pi^{(t)}(n) = \frac{(tn)^n}{n.n!}e^{-tn},$$

and $l(t)$ follows a size biased Borel law :

$$\mathbb{P}(l(t) = n) = (1-t)\frac{(tn)^n}{n!}e^{-tn}.$$

(2) An other example where calculus can be made explicitly is the gamma case when $\nu(\mathrm{d}l) = \mathbb{1}_{\{l \geq 0\}} l^{-1} e^{-l} \mathrm{d}l$. Note that $\bar{\nu}(0) = \infty$ and $\mathrm{m} = 1$. Then, for every $x \in \mathbb{R}_+$,

$$\mathbb{P}(Y_x^{(t)} \in \mathrm{d}z) = \mathbb{1}_{[-x,\infty[}(z)\Gamma(tx)^{-1}e^{-(z+x)}(z+x)^{tx-1}\mathrm{d}z,$$
$$\mathbb{P}(\overrightarrow{\tau}_x^{(t)} \in \mathrm{d}z) = \mathbb{1}_{[x,\infty[}(z)x(z\Gamma(tz))^{-1}e^{-(z-x)}(z-x)^{tz-1}\mathrm{d}z. \tag{3.12}$$

Further

$$\Pi^{(t)}(\mathrm{d}z) = (z\Gamma(tz))^{-1}e^{-z}z^{tz-1}\mathrm{d}z,$$

and

$$\mathbb{P}(l(t) \in \mathrm{d}x) = (1-t)\big(\delta_0(\mathrm{d}x) + \Gamma(tz)^{-1}e^{-x}x^{tx-1}\mathrm{d}x\big).$$

(3) For the exponential distribution $\nu(\mathrm{d}l) = \mathbb{1}_{\{l \geq 0\}} e^{-l}\mathrm{d}l$, we can get :

$$\Psi^{(t)}(\lambda) = \lambda(-1 + \frac{t}{\lambda+1}), \quad \kappa^{(t)}(\lambda) = (\lambda + t - 1 + \sqrt{(\lambda+t-1)^2 + 4\lambda})/2.$$

Finally, we specify two distributions involved in the storage of the data.

Writing $-g(t) = \gamma(t)$ (see (3.15) and (3.16)) and using the identity of fluctuation (2.18) gives an other expression for the Laplace transform of $g(t)$: For all $t \in [0, 1/\mathrm{m}[$ and $\lambda \geq 0$, we have

$$\mathbb{E}\big(\exp\big(\lambda g(t)\big)\big) = \exp\bigg(\int_0^\infty (e^{-\lambda x} - 1)x^{-1}\mathbb{P}(Y_x^{(t)} > 0)\mathrm{d}x\bigg). \tag{3.13}$$

As a consequence, we see that the law of $g(t)$ is infinitively divisible. Moreover this expression will give the generating triplet of the additive process $(g(t))_{t\in[0,1/\mathrm{m}[}$ (see Theorem 2 in Section 3.4 in [12]).

The quantity of data over 0, $R_0^{(t)}$ (see Section 3.2), is an increasing process equal to $(-I_0^{(t)})_{t\geq 0}$. By (3.15) and (2.17), its Laplace transform is then equal to

$$\lambda \longrightarrow \frac{(1-\mathrm{m}t)\lambda}{\Psi^{(t)}(\lambda)}.$$

3.4 Proofs

In this section, we provide rigorous arguments for the original results which have been stated in Section 3.3.

Proof of Proposition 3.3.1. First $\mathrm{m} < \infty$ entails that $\forall L \geq 0$, $\sum_{t_i \leq t, x_i \in [-L,L]} l_i < \infty$ a.s. and condition (2.12) is satisfied a.s.

• If $t < 1/\mathrm{m}$, then $\mathbb{E}(Y_{-1}^{(t)}) = 1 - \mathrm{m}t > 0$ and the càdlàg version of $(Y_{(-x)^-}^{(t)})_{x\geq 0}$ is a Lévy process. So we have (see [19] Corollary 2 on page 190) :

$$Y_x^{(t)} \overset{x\to-\infty}{\longrightarrow} \infty \quad \text{a.s.}$$

Then Proposition 3.2.2 ensures that for every $t < 1/\mathrm{m}$, $\mathcal{C}(t) = \{x \in \mathbb{R} :\ Y_x^{(t)} > I_x^{(t)}\} \neq \mathbb{R}$ a.s.

• If $t \geq 1/\mathrm{m}$, then $\mathbb{E}(Y_{-1}^{(t)}) \leq 0$ ensures (see [19] Corollary 2 on page 190) :

$$Y_x^{(t)} \overset{x\to-\infty}{\longrightarrow} -\infty \text{ a.s} \quad \text{or} \quad (Y_x^{(t)})_{x\leq 0} \text{ oscillates a.s in } -\infty.$$

Similarly, we get that for every $t \geq 1/\mathrm{m}$, $\mathcal{C}(t) = \mathbb{R}$ a.s. □

For the forthcoming proofs, we fix $t \in [0, 1/\mathrm{m}[$, which is omitted from the notation of processes for the sake of simplicity.

To prove the regeneration property and characterize the Laplace exponent of $\vec{\tau}$, we need to establish first a regeneration property at the right extremities of the data blocks. In that view, we consider for every $x \geq 0$, the files arrived at the left/at the right of x before time t :

$$\mathcal{P}_x := \{(t_i, x_i, l_i) : t_i \leq t,\ x_i \leq x\}, \quad \mathcal{P}^x := \{(t_i, x_i - x, l_i) : t_i \leq t,\ x_i > x\}.$$

Lemma 3.4.1. *For all $x \geq 0$, $\mathcal{P}^{d_x(\mathcal{R}(t))}$ is independent of $\mathcal{P}_{d_x(\mathcal{R}(t))}$ and distributed as $\mathcal{P}^0$.*

Proof. The simple Markov property for PPP states that for every $x \in \mathbb{R}$, $\mathcal{P}^x$ is independent of $\mathcal{P}_x$ and distributed as $\mathcal{P}^0$. Clearly this extends to simple stopping times in the filtration $\sigma\big(\mathcal{P}_x\big)_{x\in\mathbb{R}}$ and further to any stopping time in this filtration using the classical argument of approximation of stopping times by a decreasing sequence of simple stopping times (see also [73]). As $d_x(\mathcal{R}(t))$ is a stopping time in this filtration, $\mathcal{P}^{d_x(\mathcal{R}(t))}$ is independent of $\mathcal{P}_{d_x(\mathcal{R}(t))}$ and distributed as $\mathcal{P}^0$. □

Proof of Proposition 3.3.2. • The free space at the right of $d_x(\mathcal{R}(t))$ at time t is given by the point process of files arrived at the right of $d_x(\mathcal{R}(t))$ before time t. That is, there exists a measurable functional F such that for all $x \in \mathbb{R}$,

$$(\mathcal{R}(t) - d_x(\mathcal{R}(t))) \cap [0, \infty[= F\big(\mathcal{P}^{d_x(\mathcal{R}(t))}\big).$$

Similarly $\mathcal{R}(t)\cap] - \infty, x]$ is $\mathcal{P}_{d_x(\mathcal{R}(t))}$ measurable. The previous lemma ensures then that $(\mathcal{R}(t) - d_x(\mathcal{R}(t))) \cap [0, \infty[$ is independent of $\mathcal{R}(t)\cap] - \infty, x]$ and is distributed as $(\mathcal{R} - d_0(\mathcal{R}(t))) \cap [0, \infty[$.

• The stationarity of $\mathcal{C}(t)$ should be plain from the construction of the covering and the fact that the law of a PPP with intensity $\mathrm{d}x \otimes \nu$ is invariant by translation of the first coordinate. Stationarity can also be viewed as a consequence of regeneration and $\inf \mathcal{R}(t) = -\infty$ (see Remark (4.11) in [71]).

• The symmetry of $\mathcal{R}(t)^{cl}$ is a consequence of the regeneration property and stationarity (see Lemma 6.5 in [89] or Corollary (7.19) in [90]).

• As a consequence of stationarity, $\mathbb{P}(x \in \mathcal{C}(t))$ does not depend on x and is equal to $\mathbb{P}(0 \in \mathcal{C}(t))$. Following Section 2.1, we write $R_x := Y_x - I_x$ the quantity of data over x so that the quantity of data stored in $[-L, L]$ is given for every $L > 0$ by

$$\mid \mathcal{C}(t) \cap [-L, L] \mid= R_{-L} + \Big(\sum_{t_i \leq t,\ x_i \in]-L, L]} l_i\Big) - R_L.$$

By invariance of the PPP $\{(t_i, x_i, l_i) : i \in \mathbb{N}\}$ by translation of the second coordinate,

$$\mathbb{P}\big((2L)^{-1}R_L \geq \epsilon\big) = \mathbb{P}\big((2L)^{-1}R_{-L} \geq \epsilon\big) = \mathbb{P}\big((2L)^{-1}R_0 \geq \epsilon\big) \overset{L\to\infty}{\longrightarrow} 0.$$

Moreover using (2.8), $(2L)^{-1} \sum_{t_i \leq t,\ x_i \in]-L,L[} l_i \overset{L\to\infty}{\longrightarrow} \mathrm{m}t$ in probability. So

$$\mathbb{E}\big((2L)^{-1} \mid \mathcal{C}(t) \cap [-L, L] \mid \big) \overset{L\to\infty}{\longrightarrow} \mathrm{m}t$$

and we conclude with

$$\mathbb{E}\big(\mid \mathcal{C}(t) \cap [-L, L] \mid \big) = \mathbb{E}\big(\int_{-L}^{L} \mathbb{1}_{\{x \in \mathcal{C}(t)\}} \mathrm{d}x \big) = \int_{-L}^{L} \mathbb{P}(x \in \mathcal{C}(t)) \mathrm{d}x = 2L\mathbb{P}(0 \in \mathcal{C}(t)).$$

One can also give a formal argument using Theorem 1 in [89] or $\mathbb{P}(0 \in \mathcal{C}(t)) = \mathbb{P}(l(t) > 0)$ and Theorem 3.3.3. $\square$

Proof of Proposition 3.3.3. (i) By symmetry of $\mathcal{R}(t)^{cl}$, $\overrightarrow{\mathcal{R}(t)}$ and $\overleftarrow{\mathcal{R}(t)}$ are identically distributed. The regeneration property ensures that $\overrightarrow{\mathcal{R}(t)}$ is independent of $(\overleftarrow{\mathcal{R}(t)}, g(t), d(t))$. By symmetry, $\overleftarrow{\mathcal{R}(t)}$ is independent of $(g(t), d(t), \overrightarrow{\mathcal{R}(t)})$. So $\overrightarrow{\mathcal{R}(t)}$, $\overleftarrow{\mathcal{R}(t)}$ and $(g(t), d(t))$ are independent.

(ii) As $\overrightarrow{\mathcal{R}}(t)$ is a.s. the union of intervals of the form $[a, b[$, then $x \to |\mathcal{R}(t) \cap [0, x]|$ increases at $x \in \overrightarrow{\mathcal{R}}(t)$. So, for every $x \geq 0$,

$$\overrightarrow{\tau}_{|\mathcal{R}(t)\cap[0,x]|} = d_x(\mathcal{R}(t)), \qquad \overrightarrow{\tau}_x = d_{\overrightarrow{\tau}_x}(\mathcal{R}(t)) \qquad a.s.$$

So the range of $\overrightarrow{\tau}$ is equal to $\overrightarrow{\mathcal{R}(t)}$. The fact $\overrightarrow{\tau}$ is a subordinator will be proved below but could be also derived directly from the regeneration property of $\overrightarrow{\mathcal{R}}(t)$ (see [71]). Similarly the range of $\overleftarrow{\tau}$ is equal to $\overleftarrow{\mathcal{R}(t)}$.
Moreover, $\mathrm{d}Y = -1$ on $\mathcal{R}(t)$ and $Y_{a-} = Y_b$ if $[a, b[$ is an interval component of $\mathcal{C}(t)$. By integrating on $[d(t), d(t) + y]$, we have a.s for every $y \geq 0$ such that $d(t) + y \in \mathcal{R}(t)$,

$$Y_{y+d(t)} - Y_y = -|\mathcal{R}(t) \cap [d(t), d(t) + y]|.$$

Then using again the definition of $\overrightarrow{\tau}$ given in Section 3.3.1 and that $\overrightarrow{\mathcal{R}(t)}$ is the range of $\overrightarrow{\tau}$,

$$\begin{aligned} \overrightarrow{\tau}_x &= \inf\{y \geq 0 : y \in \overrightarrow{\mathcal{R}(t)},\ |\overrightarrow{\mathcal{R}(t)} \cap [0, y]| > x\} \\ &= \inf\{y \geq 0 : d(t) + y \in \mathcal{R}(t),\ |\mathcal{R}(t) \cap [d(t), d(t) + y]| > x\} \\ &= \inf\{y \geq 0 : Y_{y+d(t)} - Y_{d(t)} < -x\}. \end{aligned} \tag{3.14}$$

Moreover

$$Y_{y+d(t)} - Y_{d(t)} = -y + \sum_{\substack{(t_i,x_i,l_i)\in\mathcal{P}^{d(t)}\\ 0\le x_i\le y}} l_i,$$

and Lemma 3.4.1 entails that $\mathcal{P}^{d(t)}$ is distributed as a PPP on $[0,t]\times\mathbb{R}_+\times\mathbb{R}_+$ with intensity $\mathrm{d}s\otimes\mathrm{d}x\otimes\nu(\mathrm{d}l)$. So $(Y_{y+d(t)} - Y_{d(t)})_{y\ge 0}$ is a Lévy process with bounded variation and drift -1 which verifies condition (2.12) (use (2.10) and $-1+\mathrm{m}t<0$). Then Theorem 2.2.1 entails that $\overrightarrow{\tau}$ is a subordinator whose Laplace exponent is the inverse function of $-\Psi^{(t)}$.

As $\overleftarrow{\mathcal{R}(t)}$ is distributed as $\overrightarrow{\mathcal{R}(t)}$, $\overleftarrow{\tau}$ is distributed as $\overrightarrow{\tau}$ by definition.

(iii) We determine now the distribution of $(g(t), d(t))$ using fluctuation theory, which enables us to get identities useful for the rest of the work. We write $(\widetilde{Y}_x)_{x\ge0}$ for the càdlàg version of $(-Y_{-x})_{x\ge0}$ and

$$S(t) := \sup\{\widetilde{Y}_x, x\ge 0\} = -I_0, \qquad \gamma(t) = \inf\{x\ge 0 : \widetilde{Y}_x = S(t)\}. \tag{3.15}$$

Using (3.4) and the fact that Y has no negative jumps, we have

$$\begin{aligned} g(t) &= g_0(\mathcal{R}(t)) = \sup\{x\le 0:\ Y_x = I_x\} \\ &= \sup\{x\le 0:\ Y_{x-} = I_0\} = -\inf\{x\ge 0:\ \widetilde{Y}_x = -I_0\} \\ &= -\gamma(t). \end{aligned} \tag{3.16}$$

Using again (3.4) and the fact that $(Y_x)_{x\ge0}$ is regular for $]-\infty,0[$ (see [19] Proposition 8 on page 84), we have also a.s.

$$\begin{aligned} d(t) &= \inf\{x>0 : Y_x = I_x\} = \inf\{x>0 : Y_x = I_0\} \\ &= \inf\{x>0 : Y_x < I_0\} = \inf\{x>0 : Y_x < -S(t)\} = T_{S(t)}, \end{aligned}$$

where $(T_x)_{x\ge0}$ is distributed as $(\overrightarrow{\tau}_x)_{x\ge0}$ by (3.14) and $(T_x)_{x\ge0}$ is independent of $(S(t),\gamma(t))$ since $(Y_x)_{x\ge0}$ is independent of $(Y_x)_{x\le0}$. Then for all $\lambda,\mu\ge0$ with $\lambda\ne\mu$:

$$\begin{aligned} \mathbb{E}\big(\exp(\lambda g(t) - \mu d(t))\big) &= \mathbb{E}\big(\exp(-\lambda\gamma(t))\mathbb{E}(\exp(-\mu T_{S(t)}))\big) \\ &= \mathbb{E}\big(\exp(-\lambda\gamma(t) - \kappa^{(t)}(\mu)S(t))\big) \\ &= -[\Psi^{(t)}]'(0)\frac{\kappa^{(t)}(\lambda)-\kappa^{(t)}(\mu)}{\lambda-\mu} \quad \text{using (2.16)} \qquad (3.17) \\ &= (1-\mathrm{m}t)\frac{\kappa^{(t)}(\lambda)-\kappa^{(t)}(\mu)}{\lambda-\mu} \quad \text{using (2.10)}, \qquad (3.18) \end{aligned}$$

which gives the distributions of $d(t)$, $g(t)$ and $l(t)$ letting respectively $\lambda=0$, $\mu=0$ and $\lambda\to\mu$. Computing then the Laplace transform of $(-Ul(t),(1-U)l(t))$ where

U is a uniform random variable on $[0, 1]$ independent of $l(t)$ gives the right hand side of (3.18). So $(g(t), d(t)) = (-U'l(t), (1-U')l(t))$, where U' is a uniform random variable on $[0, 1]$ independent of $l(t)$. $\square$

Remark 3. We have proved above that $\overleftarrow{\mathcal{R}(t)}$ is distributed as $\overrightarrow{\mathcal{R}(t)}$, which entails that the last passage-time-process of the post-infimum process of $(-Y_x)_{x\geq 0}$ is distributed as the first-passage-time process of $(-Y_x)_{x\geq 0}$.
This result is also a consequence of the fact that the post-infimum process of $(-Y_x)_{x\geq 0}$ is distributed as the Lévy process $(-Y_x)_{x\geq 0}$ conditioned to stay positive [74], whose last-passage-time process is a subordinator with Laplace exponent κ (see Exercise 3 on page 213 in [19]).

Proof of Corollary 3.3.4. As $\bar{\nu}(0) < \infty$, then $\bar{\Pi}(0) = t\bar{\nu}(0) < \infty$ (see (3.8)). So $\overrightarrow{\tau}$ is the sum of a drift and a compound Poisson process. That is, there exists a Poisson process $(N_x)_{x\geq 0}$ of intensity $t\bar{\nu}(0)$ and a sequence $(X_i)_{i\in\mathbb{N}}$ of iid variables of law $\nu/\bar{\nu}(0)$ independent of $(N_x)_{x\geq 0}$ such that

$$\overrightarrow{\tau}_x = x + \sum_{i=1}^{N_x} X_i, \qquad x \geq 0.$$

As $\overrightarrow{\mathcal{R}(t)}$ is the range of $\overrightarrow{\tau}$, the number of data blocks of $\mathcal{C}(t)$ between $d(t)$ and $d(t) + \overrightarrow{\tau}_x$ is equal to the number of jumps of $\overrightarrow{\tau}$ before x, that is N_x. Thus,

$$\frac{\text{number of data blocks in } [d(t), d(t) + \overrightarrow{\tau}_x]}{\overrightarrow{\tau}_x} = \frac{N_x}{\overrightarrow{\tau}_x} \overset{x\to\infty}{\longrightarrow} \frac{\mathbb{E}(N_1)}{\mathbb{E}(\overrightarrow{\tau}_1)} = t\bar{\nu}(0)(1 - \mathrm{m}t) \quad \text{a.s.}$$

by the law of large numbers (see [19] on page 92). This completes the proof. $\square$

Chapter 4

Asymptotic regimes

In this Chapter, we determine the asymptotic behavior of $\mathcal{C}(t)$ (the set of occupied locations at time t) at saturation of the hardware, i.e as $t \to 1/\mathrm{m}$ (Section 4.1). Recall that $\mathcal{C}(t)$ tends to $\mathbb{R}$ as $t \to 1/\mathrm{m}$ and it is more convenient to consider $\mathcal{R}(t)$, the complementary set of $\mathcal{C}(t)$. We also give in Section 4.2 the asymptotic behavior of $\mathcal{C}(t) \cap [0, x]$ when x tends to infinity and t tends to $1/\mathrm{m}$. As expected, we recover the phase transition of the largest block of the hardware observed by Chassaing and Louchard in [28].

We first give several definitions which will be useful for the study of the asymptotic regimes. Following the notation in [23], we say that $\nu \in \mathcal{D}_{2+}$ if ν has a finite second moment $\mathrm{m}_2 := \int_0^\infty l^2 \nu(\mathrm{d}l)$. For $\alpha \in]1, 2]$, we say that $\nu \in \mathcal{D}_\alpha$ whenever

$$\exists C > 0 \ \text{ such that } \ \bar{\nu}(x) \overset{x \to \infty}{\sim} C x^{-\alpha}.$$

Then, for $\alpha \in]1, 2[$, we put :

$$C_\alpha := \left(\frac{C\Gamma(2-\alpha)}{\mathrm{m}(\alpha - 1)} \right)^{1/\alpha}.$$

We denote by $(B_z)_{z \in \mathbb{R}}$ a two-sided Brownian motion, i.e. $(B_x)_{x \geq 0}$ and $(B_{-x})_{x \geq 0}$ are independent standard Brownian motions. For $\alpha \in]1, 2[$, we denote by $(\sigma_z^{(\alpha)})_{z \in \mathbb{R}}$ a càdlàg process with independent and stationary increments such that $(\sigma_x^{(\alpha)})_{x \geq 0}$ is a standard spectrally positive stable Lévy process with index α :

$$\forall\, x \geq 0, \lambda \geq 0, \quad \mathbb{E}\big(\exp(-\lambda \sigma_x^{(\alpha)})\big) = \exp(x \lambda^\alpha).$$

Finally, for all $\lambda \geq 0$ and $\alpha \in]1, 2[$, we introduce the following processes indexed by $z \in \mathbb{R}$

$$Y_z^{2+,\lambda} = -\lambda z + \sqrt{\mathrm{m}_2/\mathrm{m}} B_z, \qquad Y_z^{2,\lambda} = -\lambda z + \sqrt{C/\mathrm{m}} B_z, \qquad Y_z^{\alpha,\lambda} = -\lambda z + C_\alpha \sigma_z^{(\alpha)},$$

and their infimum process defined $I_x^{\alpha,\lambda} := \inf\{Y_y^{\alpha,\lambda} : y \leq x\}$ for $x \in \mathbb{R}$.

Topology of Matheron. If I is a closed interval of $\mathbb{R}$, we denote by $\mathcal{H}(I)$ the space of closed subsets of I. For all $x, y \in \mathbb{R}$ and $A \subset \mathbb{R}$ we define

$$d(x,y) = 1 - e^{-|x-y|}, \qquad d(x,A) = \inf\{d(x,y) : y \in A\},$$

and we endow $\mathcal{H}(I)$ with the Hausdorff distance d_H defined for all $A, B \in \mathcal{H}(I)$ by :

$$d_H(A,B) = \max\big(\sup_{x\in A} d(x,B), \sup_{x\in B} d(x,A)\big).$$

The topology induced by this distance is the topology of Matheron [72] : a sequence $\mathcal{R}_n$ in $\mathcal{H}(I)$ converges to $\mathcal{R}$ iff for each open set G and each compact K,

$$\begin{aligned} &\mathcal{R} \cap G \neq \varnothing \quad \text{implies} \quad \mathcal{R}_n \cap G \neq \varnothing \quad \text{for } n \text{ large enough},\\ &\mathcal{R} \cap K = \varnothing \quad \text{implies} \quad \mathcal{R}_n \cap K = \varnothing \quad \text{for } n \text{ large enough}. \end{aligned}$$

It is also the topology induced by the Hausdorff metric on a compact set using $\arctan(\mathcal{R} \cup \{-\infty, \infty\})$ or the Skorokhod metric using the class of 'descending saw-tooth functions' (see [72] and [40] for details).

4.1 Asymptotics at saturation of the hardware

We focus now on the asymptotic behavior of $\mathcal{R}(t)$ when t tends to $1/\mathrm{m}$, that is when the hardware is becoming full. First, note that if ν has a finite second moment, then

$$\mathbb{E}\big(l(t)\big) = \frac{\int_0^\infty l^2 \nu(\mathrm{d}l)}{(1-\mathrm{m}t)^2}.$$

Thus we may expect that if ν has a finite second moment, then $(1-\mathrm{m}t)^2 l(t)$ should converge in distribution as t tends to $1/\mathrm{m}$. Indeed, in the particular case $\nu = \delta_1$ or in the conditions of Corollary 2.4 in [21], we have an expression of $\Pi^{(t)}(\mathrm{d}x)$ and we can prove that $(1-\mathrm{m}t)^2 l(t)$ does converge in distribution to a gamma variable.

More generally, we shall prove that the rescaled free space $(1-\mathrm{m}t)^2\mathcal{R}(t)$ converges in distribution as t tends to $1/\mathrm{m}$. In that view, we need to prove that the process $(Y^{(t)}_{(1-\mathrm{m}t)^{-2}x})_{x\in\mathbb{R}}$ converges after suitable rescaling to a random process. Thanks to (3.4), $(1-\mathrm{m}t)^2\mathcal{R}(t)$ should then converge to the set of points where this limiting process coincides with its infimum process. We shall also handle the case where ν has an infinite second moment and find the correct normalization, which depends on the tail of ν. Proofs are close to proofs of the next section and they are made simultaneously in the last section of this paper.

In queuing systems, asymptotics at saturation are known as heavy traffic approximation ($\rho = t\mathrm{m} \to 1$), which depend similarly on the tail of ν. And for ν

finite, results given here could be directly derived from results in queuing theory (See III.7.2 in [30] or [62] if ν has a second moment order and [27] for heavy tail of ν). The main difference is that ν can be infinite in this paper. Then the busy cycle is not defined and we consider the whole random set of occupied locations. Moreover, as explained below, asymptotics of $\mathcal{R}(t)$ can not be directly derived from asymptotics of Y or the workload R.

We introduce now the following functions defined for every $t \in [0, 1/\mathrm{m}[$ and $\alpha \in (1,2)$ by

$$\epsilon_{2+}(t) = (1-\mathrm{m}t)^2, \qquad \epsilon_2(t) = 2\frac{(1-\mathrm{m}t)^2}{-\log((1-\mathrm{m}t))}, \qquad \epsilon_\alpha(t) = (1-\mathrm{m}t)^{\frac{\alpha}{\alpha-1}}.$$

We have then the following weak convergence result for the Hausdorff metric defined above (see Section 3.4 for the proof).

Theorem 4.1.1. *If $\nu \in \mathcal{D}_\alpha$ ($\alpha \in (1,2] \cup \{2+\}$), then $\epsilon_\alpha(t).\mathcal{R}(t)^{cl}$ converges weakly in $\mathcal{H}(\mathbb{R})$ as t tends to $1/\mathrm{m}$ to $\{x \in \mathbb{R} : Y_x^{\alpha,1} = I_x^{\alpha,1}\}^{cl}$.*

First we prove the convergence of the Laplace exponent $\Psi^{(t)}$ after suitable rescaling as t tends to $1/\mathrm{m}$, which ensures the convergence of the Lévy process $Y^{(t)}$ after suitable rescaling (see Lemma 4.3.1). These convergences will not a priori entail the convergence of the random set $\epsilon_\alpha(t).\mathcal{R}^{cl}(t)$ since they do not entail the convergence of excursions. Nevertheless, they will entail the convergence of $\kappa^{(t)}$ since $\kappa^{(t)} \circ (-\Psi^{(t)}) = \mathrm{Id}$ (Lemma 4.3.2). Then we get the convergence of $\tau^{(t)}$ as t tends to infinity and thus of its range $\epsilon_\alpha(t).\mathcal{R}^{cl}(t)$.

Remark 4. More generally, as in queuing theory and [23], we can generalize these results for regularly varying functions $\bar{\nu}$. If $\bar{\nu}$ is regularly varying at infinity with index $-\alpha \in (-1,-2)$, then we have the following weak convergence in $\mathcal{H}(\mathbb{R})$:

$$z^{-1}\mathcal{R}((1-z\bar{\nu}(z))/\mathrm{m})^{cl} \overset{z\to\infty}{\Longrightarrow} \{x \in \mathbb{R} : Y_x^{\alpha,1} = I_x^{\alpha,1}\}^{cl} \quad \text{with } C = 1.$$

For instance, the case $\bar{\nu}(x) \overset{x\to\infty}{\sim} cx^{-\alpha}\log(x)^\beta$ with $(\alpha,\beta,c) \in]1,2[\times\mathbb{R}\times\mathbb{R}_+^*$ leads to

$$\big((1-\mathrm{m}t)\log(1/(1-\mathrm{m}t))^{-\beta}\big)^{\frac{1}{\alpha-1}}\mathcal{R}(t)^{cl} \overset{t\to 1/\mathrm{m}}{\Longrightarrow} \{x \in \mathbb{R} : Y_x^{\alpha,1} = I_x^{\alpha,1}\}^{cl},$$

with $C = c/(\alpha-1)^\beta$. If $\bar{\nu}$ is regularly varying at infinity with index -2, there are many cases to consider.

We get then the asymptotic of $(g(t), d(t))$:

Corollary 4.1.2. *If $\nu \in \mathcal{D}_\alpha$ ($\alpha \in (1,2] \cup \{2+\}$), then $\epsilon_\alpha(t).(g(t),d(t))$ converges weakly as t tends to $1/\mathrm{m}$ to $(\sup\{x \le 0 : Y_x^{\alpha,1} = I_0^{\alpha,1}\}, \inf\{x \ge 0 : Y_x^{\alpha,1} = I_0^{\alpha,1}\})$. If $\nu \in \mathcal{D}_{2+}$ (resp. $\mathcal{D}_2$), $\epsilon_\alpha(t).l(t)$ converges weakly to a gamma variable with parameter $(1/2, \mathrm{m}/(4\mathrm{m}_2))$ (resp. $(1/2, \mathrm{m}/4)$).*

Remark 5. The density of data blocks of size $\mathrm{d}x$ in $\epsilon_\alpha(t).\mathcal{R}(t)^{cl}$ is equal to $\frac{\mathrm{m}t}{1-\mathrm{m}t}\Pi^{(t)}(\mathrm{d}x)$. By the previous theorem or corollary, this density converges weakly as t tends to $1/\mathrm{m}$ to the density of data block of size $\mathrm{d}x$ of the limit covering $\{x \in \mathbb{R} : Y_x^{\alpha,1} = I_x^{\alpha,1}\}^{cl}$. This limit density, denoted by $\Pi^{\alpha,1}(\mathrm{d}x)$, can be computed explicitly in the cases $\nu \in \mathcal{D}_\alpha$ ($\alpha \in \{2, 2+\}$), thanks to the last corollary :

$$\Pi^{2+,1}(\mathrm{d}x) = \sqrt{\frac{\mathrm{m}}{4\pi\mathrm{m}_2 x^3}}\exp\big(-\frac{\mathrm{m}}{4\mathrm{m}_2}x\big), \qquad \Pi^{2,1}(\mathrm{d}x) = \sqrt{\frac{\mathrm{m}}{4\pi x^3}}\exp\big(-\frac{\mathrm{m}}{4}x\big).$$

Note that is also the Lévy measure of the limit covering $\{x \in \mathbb{R} : Y_x^{\alpha,1} = I_x^{\alpha,1}\}^{cl}$.

4.2 Asymptotic regime on a large part of the hardware

Here we look at the set of occupied locations $\mathcal{C}(t)$ in a window of size x. We consider the asymptotics of $\mathcal{C}(t) \cap [0, x]$ when x tends to infinity at saturation time. As far as we know, results given here are new even when ν is finite. We introduce the following functions defined for all $x \in \mathbb{R}_+^*$ and $\alpha \in (1, 2)$ by

$$f_{2+}(x) = 1/\sqrt{x}, \qquad f_2(x) = \sqrt{\log(x)/x}, \qquad f_\alpha(x) = x^{1/\alpha-1}.$$

And we have the following asymptotic regime (see Section 3.4 for the proof).

Theorem 4.2.1. *If $\nu \in \mathcal{D}_\alpha$ ($\alpha \in (1, 2] \cup \{2+\}$), x tends to infinity and t to $1/\mathrm{m}$ such that $1 - \mathrm{m}t \sim \lambda f_\alpha(x)$ with $\lambda > 0$, then $x^{-1}(\mathcal{R}(t)^{cl} \cap [0, x])$ converges weakly in $\mathcal{H}([0, 1])$ to $\{z \in [0, 1] : Y_z^{\alpha,\lambda} = I_z^{\alpha,\lambda}\}^{cl}$.*

Thus as in [28], we observe a phase transition of the size of largest block of data in $[0, x]$ as $x \to \infty$ according to the rate of filling of the hardware. More precisely, denoting $B_1(x, t) =| I_1(x, t) |$ where $(I_j(x, t))_{j\geq 1}$ is the sequence of component intervals of $\mathcal{C}(t) \cap [0, x]$ ranked by decreasing order of size, we have :

Corollary 4.2.2. *Let $\nu \in \mathcal{D}_\alpha$ ($\alpha \in (1, 2] \cup \{2+\}$), x tend to infinity and t to $1/\mathrm{m}$:*
- If $1 - \mathrm{m}t \sim \lambda f_\alpha(x)$ with $\lambda > 0$, then $B_1(x, t)/x$ converges in distribution to the largest length of excursion of $(Y_z^{\alpha,\lambda} - I_z^{\alpha,\lambda})_{z\in[0,1]}$.
- If $f_\alpha(x) = o(1 - \mathrm{m}t)$, then $B_1(x, t)/x \xrightarrow{\mathbb{P}} 0$.
- If $1 - \mathrm{m}t = o(f_\alpha(x))$, then $B_1(x, t)/x \xrightarrow{\mathbb{P}} 1$.

The phase transition occurs at time t such that $1 - \mathrm{m}t \sim \lambda f_\alpha(x)$ with $\lambda > 0$. The more data arrive in small files (i.e. the faster $\bar{\nu}(x)$ tends to zero as x tends to infinity), the later the phase transition occurs. In [28, 23], the hardware is a circle and processes required for asymptotics are the bridges of the processes used here. A consequence is that in our model, $B_1(t, x)/x$ tends to one with a positive probability at phase transition, which is not the case for the parking problem in [28, 23]. More precisely, denoting by $B_{\alpha,\lambda}$ the law of the largest length of excursion of $(Y_x^{\alpha,\lambda} - I_x^{\alpha,\lambda})_{x\in[0,1]}$, we have :

$$\forall\ (\lambda,\alpha) \in \mathbb{R}_+^* \times]1,2[\cup\{2+\}, \qquad \mathbb{P}(B_{\alpha,\lambda} = 1) > 0.$$

4.3 Proofs

Proofs of Theorem 4.1.1 and Theorem 4.2.1 are close and made simultaneously. For that purpose, we introduce now $\Psi^{\alpha,\lambda}$ the Laplace exponent (see (2.7)) of $Y^{\alpha,\lambda}$ given for $y \geq 0$, $\lambda \geq 0$ and $\alpha \in (1,2)$ by

$$\Psi^{2+,\lambda}(y) = -\lambda y - \frac{\mathrm{m}_2}{\mathrm{m}}\frac{y^2}{2}, \qquad \Psi^{2,\lambda}(y) = -\lambda y - \frac{C}{\mathrm{m}}\frac{y^2}{2}, \qquad \Psi^{\alpha,\lambda}(y) = -\lambda y - (C_\alpha y)^\alpha.$$

We denote by $\mathbb{D}$ the space of càdlàg function from $\mathbb{R}_+$ to $\mathbb{R}$ which we endow with the Skorokhod topology (see [54] on page 292). First, we prove the weak convergence of $Y^{(t)}$ after suitable rescaling.

Lemma 4.3.1. *If $\nu \in \mathcal{D}_\alpha$ ($\alpha \in (1,2] \cup \{2+\}$), then for all $y \geq 0$ and $\lambda > 0$:*

$$\epsilon_\alpha(t)^{-1}\Psi^{(t)}(\epsilon_\alpha(t)(1-\mathrm{m}t)^{-1}y) \overset{t\to 1/\mathrm{m}}{\longrightarrow} \Psi^{\alpha,1}(y),$$
$$x\Psi^{((1-\lambda f_\alpha(x))/\mathrm{m})}((xf_\alpha(x))^{-1}y) \overset{x\to\infty}{\longrightarrow} \Psi^{\alpha,\lambda}(y),$$

which entail the following weak convergences of processes in $\mathbb{D}$:

$$\left(\epsilon_\alpha(t)(1-\mathrm{m}t)^{-1}Y^{(t)}_{\epsilon_\alpha(t)^{-1}y}\right)_{y\geq 0} \overset{t\to 1/\mathrm{m}}{\Longrightarrow} (Y_y^{\alpha,1})_{y\geq 0},$$

$$\left((xf_\alpha(x))^{-1}Y^{((1-\lambda f_\alpha(x))/\mathrm{m})}_{xy}\right)_{y\geq 0} \overset{x\to\infty}{\Longrightarrow} (Y_y^{\alpha,\lambda})_{y\geq 0}.$$

Remark 6. If $\bar{\nu}$ is regularly varying at infinity with index $-\alpha \in (-1,-2)$, then $\bar{\nu}(x)^{-1}\Psi^{((1-\lambda x\bar{\nu}(x))/\mathrm{m})}(x^{-1}y)$ converges to $\Psi^{\alpha,\lambda}(y)$ as x tends to infinity.

Proof of Lemma 4.3.1. Using (2.9), we have

$$x\Psi^{(t)}(y) = xy(\mathrm{m}t - 1 - t\int_0^\infty (1-e^{-yu})\bar{\nu}(u)\mathrm{d}u). \tag{4.1}$$

We handle now the different cases :

• Case $\nu \in \mathcal{D}_{2+}$. Using $|1-e^{-yu}|/y \leq u \quad (u \geq 0)$ and dominated convergence theorem gives :

$$\int_0^{\infty}(1-e^{-yu})\bar{\nu}(u)\mathrm{d}u \overset{y\to 0}{\sim} y\int_0^{\infty} u\bar{\nu}(u)\mathrm{d}u = \frac{y\mathrm{m}_2}{2},$$

which proves the first part of the lemma using (4.1).

• Case $\nu \in \mathcal{D}_\alpha$ with $\alpha \in (1,2)$. Using that $(u/y)^\alpha\bar{\nu}(u/y)$ is bounded, we apply dominated convergence theorem and get

$$\begin{aligned}\int_0^{\infty}(1-e^{-yu})\bar{\nu}(u)\mathrm{d}u &= y^{-1}\int_0^{\infty}(1-e^{-u})\bar{\nu}(u/y)\mathrm{d}u \\ &\overset{y\to 0}{\sim} y^{-1}C(y^{-1})^{-\alpha}\int_0^{\infty}(1-e^{-u})u^{-\alpha}\mathrm{d}u \\ &\overset{y\to 0}{\sim} C\frac{\Gamma(2-\alpha)}{\alpha-1}y^{\alpha-1},\end{aligned} \tag{4.2}$$

which proves the first part of the lemma using (4.1).

• Case ν is regularly varying at infinity with index $-\alpha \in]-1,-2[$. First,

$$\int_0^{1/\sqrt{y}}(1-e^{-yu})\bar{\nu}(u)\mathrm{d}u \leq y\int_0^{1/\sqrt{y}} u\bar{\nu}(u)\mathrm{d}u \overset{y\to 0}{\sim} y(1/\sqrt{y})^{2-\alpha} = y^{\alpha/2}.$$

Moreover for every $u>0$, $\bar{\nu}(u/y) \overset{y\to 0}{\sim} \bar{\nu}(y)u^{-\alpha}$. Let $\delta>0$ such that $-2<-\alpha-\delta<-\alpha+\delta<-1$. By Potter's theorem (page 25 in [26]) ensures that for all y small enough and u large enough,

$$\frac{\bar{\nu}(u/y)}{\bar{\nu}(1/y)} \leq 2\mathrm{max}(u^{-\alpha+\delta}, u^{-\alpha-\delta}).$$

So we can apply the dominated convergence theorem to get

$$\int_0^{1/\sqrt{y}}(1-e^{-yu})\bar{\nu}(u)\mathrm{d}u = y^{-1}\int_{\sqrt{y}}^{\infty}(1-e^{-u})\bar{\nu}(u/y)\mathrm{d}u \overset{y\to 0}{\sim} \frac{\Gamma(2-\alpha)}{\alpha-1}y^{-1}\bar{\nu}(1/y).$$

As $y^{\alpha/2} = o(y^{-1}\bar{\nu}(1/y)) \quad (y\to 0)$, we can complete the proof with

$$\int_0^{\infty}(1-e^{-yu})\bar{\nu}(u)\mathrm{d}u \overset{y\to 0}{\sim} \frac{\Gamma(2-\alpha)}{\alpha-1}y^{-1}\bar{\nu}(1/y).$$

• Case $\nu \in \mathcal{D}_2$. We split the integral. First, we have

$$\int_0^{1/\sqrt{y}}(1-e^{-yu})\bar{\nu}(u)\mathrm{d}u \overset{y\to 0}{\sim} y\int_0^{1/\sqrt{y}} u\bar{\nu}(u)\mathrm{d}u \overset{y\to 0}{\sim} Cy\mathrm{log}(1/y)/2.$$

since $\int_0^{1/\sqrt{y}}(1-e^{-yu}+yu)\bar{\nu}(u)\mathrm{d}u = o(y\int_0^{1/\sqrt{y}} u\bar{\nu}(u)\mathrm{d}u)$. Moreover,

$$\begin{aligned}\int_{1/\sqrt{y}}^{\infty}(1-e^{-yu})\bar{\nu}(u)\mathrm{d}u \quad &= \quad y^{-1}\int_{\sqrt{y}}^{\infty}(1-e^{-u})\bar{\nu}(u/y)\mathrm{d}u \\ &\overset{y\to 0}{\sim} \quad Cy\int_{\sqrt{y}}^{\infty}(1-e^{-u})u^{-2}\mathrm{d}u \quad \text{using } \nu\in\mathcal{D}_2 \\ &\overset{y\to 0}{\sim} \quad Cy\int_{\sqrt{y}}^{1}u^{-1}\mathrm{d}u \;=\; Cy\log(1/y)/2\end{aligned}$$

Then

$$\int_0^{\infty}(1-e^{-yu})\bar{\nu}(u)\mathrm{d}u \;\overset{y\to\infty}{\sim}\; Cy\log(1/y),$$

which proves the first part of the lemma using (4.1).

These convergences ensure the convergence of the finite-dimensional distributions of the processes. The weak convergence in $\mathbb{D}$, which is the second part of the lemma, follows from Theorem 13.17 in [55]. $\square$

In the spirit of Section 3.3.1, we introduce the expected limit set, that is the free space of the covering associated with $Y^{\alpha,\lambda}$, and the extremities of the block containing 0.

$$\mathcal{R}(\alpha,\lambda) := \{x\in\mathbb{R} : Y_x^{\alpha,\lambda} = I_x^{\alpha,\lambda}\},$$
$$g(\alpha,\lambda) := g_0(\mathcal{R}(\alpha,\lambda)), \qquad d(\alpha,\lambda) := d_0(\mathcal{R}(\alpha,\lambda)).$$

We have the following analog of Proposition 3.3.3. $\overrightarrow{\mathcal{R}(\alpha,\lambda)}$ and $\overleftarrow{\mathcal{R}(\alpha,\lambda)}$ are independent, identically distributed and independent of $(g(\alpha,\lambda), d(\alpha,\lambda))$. Moreover $\overrightarrow{\mathcal{R}(\alpha,\lambda)}$ and $\overleftarrow{\mathcal{R}(\alpha,\lambda)}$ are respectively the range of the subordinators $\overrightarrow{\tau}^{\alpha,\lambda}$ and $\overleftarrow{\tau}^{\alpha,\lambda}$, whose Laplace exponent $\kappa^{\alpha,\lambda}$ is the inverse function of $-\Psi^{\alpha,\lambda}$. Finally, using $\left[\Psi^{\alpha,\lambda}\right]'(0) = -\lambda$, the counterpart of (3.17) gives for $\rho,\mu\geq 0$ and $\rho\neq\mu$:

$$\mathbb{E}\big(\exp(\rho g(\alpha,\lambda) - \mu d(\alpha,\lambda))\big) = \lambda\frac{\kappa^{\alpha,\lambda}(\rho)-\kappa^{\alpha,\lambda}(\mu)}{\rho-\mu}. \tag{4.3}$$

The proof of these results follow the proof of Proposition 3.3.3, except for two points :

1) We cannot use the point process of files to prove the stationarity and regeneration property of $\mathcal{R}(\alpha,\lambda)$ and we must use the process $Y^{\alpha,\lambda}$ instead. The stationarity is a direct consequence of the stationarity of $\big(Y_x^{\alpha,\lambda} - I_x^{\alpha,\lambda}\big)_{x\in\mathbb{R}}$. The

regeneration property is a consequence of the counterpart of Lemma 3.4.1 which can be stated as follows. For all $x \in \mathbb{R}$,

$$\big(Y^{\alpha,\lambda}_{d_x(\mathcal{R}(\alpha,\lambda))+y} - Y^{\alpha,\lambda}_{d_x(\mathcal{R}(\alpha,\lambda))}\big)_{y\geq 0} \text{ is independent of } \big(Y^{\alpha,\lambda}_{d_x(\mathcal{R}(\alpha,\lambda))-y} - Y^{\alpha,\lambda}_{d_x(\mathcal{R}(\alpha,\lambda))}\big)_{y\geq 0}$$

and distributed as $\big(Y^{\alpha,\lambda}_y\big)_{y\geq 0}$. As Lemma 3.4.1, this property is an extension to the stopping time $d_x(\mathcal{R}(\alpha,\lambda))$ of the following obvious result : $\big(Y^{\alpha,\lambda}_{x+y} - Y^{\alpha,\lambda}_x\big)_{y\geq 0}$ is independent of $\big(Y^{\alpha,\lambda}_{x-y} - Y^{\alpha,\lambda}_x\big)_{y\geq 0}$ and distributed as $\big(Y^{\alpha,\lambda}_y\big)_{y\geq 0}$.

2) It is convenient to define directly $(\overrightarrow{\tau}^{\alpha,\lambda}_x)_{x\geq 0}$ by

$$\overrightarrow{\tau}^{\alpha,\lambda}_x := \inf\{y \geq 0 : Y^{\alpha,\lambda}_{d(\alpha,\lambda)+y} - Y^{\alpha,\lambda}_{d(\alpha,\lambda)} < -x\}.$$

For $\lambda > 0$, $[\Psi^{\alpha,\lambda}]'(0) = -\lambda < 0$ so we can apply Theorem 2.2.1 and $\overrightarrow{\tau}^{\alpha,\lambda}$ is a subordinator whose Laplace $\kappa^{\alpha,\lambda}$ is the inverse function of $-\Psi^{\alpha,\lambda}$. Moreover its range is a.s. equal to $\overrightarrow{\mathcal{R}(\alpha,\lambda)}$, since the Lévy process $(Y^{\alpha,\lambda}_{d(\alpha,\lambda)+y} - Y^{\alpha,\lambda}_{d(\alpha,\lambda)})_{y\geq 0}$ is regular for $]-\infty, 0[$ (Proposition 8 on page 84 in [19]).

To prove Theorem 4.1.1 and Theorem 4.2.1, we need a final lemma, which states the convergence of the Laplace exponent of $\overrightarrow{\mathcal{R}(t)}$.

Lemma 4.3.2. *If $\nu \in \mathcal{D}_\alpha$ ($\alpha \in (1,2] \cup \{2+\}$), then for all $z \geq 0$ and $\lambda > 0$,*

$$(1-\mathrm{m}t)\epsilon_\alpha(t)^{-1}\kappa^{(t)}(\epsilon_\alpha(t)z) \overset{t\to 1/\mathrm{m}}{\longrightarrow} \kappa^{\alpha,1}(z),$$
$$x f_\alpha(x)\kappa^{((1-\lambda f_\alpha(x))/\mathrm{m})}(x^{-1}z) \overset{x\to\infty}{\longrightarrow} \kappa^{\alpha,\lambda}(z).$$

Remark 7. If $\bar{\nu}$ is regularly varying at infinity of index $-\alpha \in (-1,-2)$, we have similarly

$$\bar{\nu}(x)^{-1}\kappa^{((1-\lambda x\bar{\nu}(x))/\mathrm{m})}(x^{-1}z) \overset{x\to\infty}{\longrightarrow} \kappa^{\alpha,\lambda}(z).$$

Proof. First we prove that

$$\alpha(t) \overset{t\to 1/\mathrm{m}}{\sim} \beta(t) \quad \Rightarrow \quad \kappa^{(t)}(\alpha(t)) \overset{t\to 1/\mathrm{m}}{\sim} \kappa^{(t)}(\beta(t)). \tag{4.4}$$

Indeed the function $u \in \mathbb{R}_+^* \mapsto \frac{1-e^{-u}}{u}$ decreases so for all $x \geq 0$ and $u, v > 0$, we have :

$$\min(\frac{u}{v}, 1) \leq \frac{1-e^{-ux}}{1-e^{-vx}} \leq \max(\frac{u}{v}, 1).$$

This gives

$$\min(\frac{\alpha(t)}{\beta(t)}, 1) \leq \frac{\int_0^\infty (1-e^{-\alpha(t)x})\Pi^{(t)}(\mathrm{d}x)}{\int_0^\infty (1-e^{-\beta(t)x})\Pi^{(t)}(\mathrm{d}x)} \leq \max(\frac{\alpha(t)}{\beta(t)}, 1),$$

and proves (4.4) recalling (3.7).

Then the first part of Lemma 4.3.1 and the identity $\kappa^{(t)} \circ (-\Psi^{(t)}) = \mathrm{Id}$ give the first part of Lemma 4.3.2. Indeed for every $y \geq 0$, $\Psi^{(t)}(\epsilon_\alpha(t)(1-\mathrm{m}t)^{-1}y) \overset{t\to 1/\mathrm{m}}{\sim} \epsilon_\alpha(t)\Psi^{\alpha,1}(y)$. So (8.14) entails

$$\epsilon_\alpha(t)(1-\mathrm{m}t)^{-1}y \overset{t\to 1/\mathrm{m}}{\sim} \kappa^{(t)}(-\epsilon_\alpha(t)\Psi^{\alpha,1}(y)).$$

Put $y = \kappa^{\alpha,1}(z)$ to get the first limit of the lemma and follow the same way to get the second one. □

Proof of Theorem 4.1.1. First, by (3.18), we have

$$\mathbb{E}\big(\exp(\rho\epsilon_\alpha(t)g(t) - \mu\epsilon_\alpha(t)d(t))\big) = (1-\mathrm{m}t)\frac{\kappa^{(t)}(\epsilon_\alpha(t)\rho) - \kappa^{(t)}(\epsilon_\alpha(t)\mu)}{\epsilon_\alpha(t)(\rho-\mu)}.$$

Letting $t \to 1/\mathrm{m}$ using Lemma 4.3.2 gives the right hand side of (4.3). Then that $\epsilon_\alpha(t).(g(t), d(t))$ converges weakly as t tends to $1/\mathrm{m}$ to $(g(\alpha,1), d(\alpha,1))$.

Moreover $\epsilon_\alpha(t)\overrightarrow{\mathcal{R}(t)}^{cl}$ (resp. $\epsilon_\alpha(t)\overleftarrow{\mathcal{R}(t)}^{cl}$) converges weakly in $\mathcal{H}(\mathbb{R}_+)$ as t tends to $1/\mathrm{m}$ to $\overrightarrow{\mathcal{R}(\alpha,1)}^{cl}$ (resp. $\overleftarrow{\mathcal{R}(\alpha,1)}^{cl}$). Indeed, by Proposition (3.9) in [40], this is a consequence of the convergence of the Laplace exponent of $\epsilon_\alpha(t)\overrightarrow{\mathcal{R}(t)}$ given by Lemma 4.3.2. Informally, $\epsilon_\alpha(t)\overrightarrow{\mathcal{R}(t)}^{cl}$ is the range of $\big(\epsilon_\alpha(t)\overrightarrow{\tau}^{(t)}_{(1-\mathrm{m}t)\epsilon_\alpha(t)^{-1}z}\big)_{z\geq 0}$ whose convergence in $\mathbb{D}$ follows from Lemma 4.3.2.

We can now prove the theorem. We know from (3.6) that

$$\epsilon_\alpha(t)\mathcal{R}(t) = \epsilon_\alpha(t).(d(t) + \overrightarrow{\mathcal{R}(t)}) \ \sqcup \ (\epsilon_\alpha(t).\widetilde{(-g(t) + \overleftarrow{\mathcal{R}(t)})})$$

where $\epsilon_\alpha(t)\overleftarrow{\mathcal{R}(t)}$, $\epsilon_\alpha(t)(-g(t), d(t))$ and $\epsilon_\alpha(t)\overrightarrow{\mathcal{R}(t)}$ are independent by Proposition 3.3.3. Similarly

$$\mathcal{R}(\alpha,1) = (d(\alpha,1) + \overrightarrow{\mathcal{R}(\alpha,1)}) \ \sqcup \ \widetilde{(-g(\alpha,1) + \overleftarrow{\mathcal{R}(\alpha,1)})}$$

where $\overleftarrow{\mathcal{R}(\alpha,1)}$, $(-g(\alpha,1), d(\alpha,1))$ and $\overrightarrow{\mathcal{R}(\alpha,1)}$ are independent. As remarked above, we have also the following weak convergences as t tends to $1/\mathrm{m}$:

$$\epsilon_\alpha(t)\overleftarrow{\mathcal{R}(t)}^{cl} \Rightarrow \overleftarrow{\mathcal{R}(\alpha,1)}^{cl}, \qquad \epsilon_\alpha(t)\overrightarrow{\mathcal{R}(t)}^{cl} \Rightarrow \overrightarrow{\mathcal{R}(\alpha,1)}^{cl},$$

$$\epsilon_\alpha(t)(-g(t), d(t)) \Rightarrow (-g(\alpha,1), d(\alpha,1)).$$

So $\epsilon_\alpha(t)\mathcal{R}(t)^{cl}$ converges weakly to $\mathcal{R}(\alpha,1)^{cl}$ in $\mathcal{H}(\mathbb{R})$ as t tends to $1/\mathrm{m}$. □

Proof of Corollary 4.1.2. The first result is a direct consequence of Theorem 4.1.1. We have then

$$\epsilon_\alpha(t)l(t) \overset{t\to 1/\mathrm{m}}{\Longrightarrow} d(\alpha,1)-g(\alpha,1).$$

Moreover, as $\kappa^{2+,1}\circ(-\Psi^{2+,1}) = \mathrm{Id}$, we can compute $\kappa^{2+,1}$ and (4.3) gives

$$\begin{aligned}\mathbb{E}\big(\exp(-\mu(d(2+,1))-g(2+,1))\big) &= (\kappa^{2+,1})'(\mu)\\ &= \Big(\frac{1+\sqrt{1+2\frac{\mathrm{m}_2}{\mathrm{m}}\mu}}{\frac{\mathrm{m}_2}{\mathrm{m}}}\Big)'(\mu)\\ &= \frac{1}{\sqrt{-1+2\frac{\mathrm{m}_2}{\mathrm{m}}\mu}}.\end{aligned}$$

So, by identification of Laplace transform, $d(\alpha,1)-g(\alpha,1)$ is a gamma variable of parameter $(1/2, \mathrm{m}/(4\mathrm{m}_2))$ and we get the result. The argument is similar in the case $\alpha = 2$. □

Proof of Theorem 4.2.1. The argument is similar to that of the proof of Theorem 4.1.1 using the others limits of Lemma 4.3.2. We get that if $x\to\infty$ and $1-\mathrm{m}t \sim \lambda f_\alpha(x)$ with $\lambda>0$, then $x^{-1}\mathcal{R}(t)$ converges weakly in $\mathcal{H}(\mathbb{R})$ to $\{x\in\mathbb{R} : Y_x^{\alpha,\lambda} = I_x^{\alpha,\lambda}\}^{cl}$. The theorem follows by restriction to $[0,1]$. □

To prove the corollary of Theorem 4.2.1, we need the following result.

Lemma 4.3.3. *The largest length of excursion of $(Y_x^{\alpha,\lambda}-I_x^{\alpha,\lambda})_{x\in[0,1]}$, denoted by $B_{\alpha,\lambda}$, converges in probability to 0 as λ tends to infinity and to 1 as λ tends to 0.*

Proof. • Let $0\le a<b\le 1$. Note that for all $\lambda'\ge 1$ and $x\ge 0$, $Y_x^{\alpha,\lambda'}-Y_x^{\alpha,1} = (1-\lambda')x$ ensures that $I_x^{\alpha,\lambda'}-I_x^{\alpha,1}\ge(1-\lambda')x$. Then,

$$Y^{\alpha,\lambda'}_{a+2\frac{b-a}{3}} - I^{\alpha,\lambda'}_{a+\frac{b-a}{3}} \le Y^{\alpha,1}_{a+2\frac{b-a}{3}} - I^{\alpha,1}_{a+\frac{b-a}{3}} + (1-\lambda')\frac{b-a}{3}.$$

So a.s there exists λ' such that

$$Y^{\alpha,\lambda'}_{a+2\frac{b-a}{3}} < I^{\alpha,\lambda'}_{a+\frac{b-a}{3}}.$$

As $Y^{\alpha,\lambda'}$ has no negative jumps, it reaches its infimum on $]-\infty, 2(b-a)/3]$ in a point $c\in[a+(b-a)/3, a+2(b-a)/3]$. Then a.s there exists $c\in[a+(b-a)/3, a+2(b-a)/3]$ and $\lambda'>0$ such that $c\in\mathcal{R}(\alpha,\lambda')$, which entails that c does not belong to the interior of $B_{\alpha,\lambda'}$. Adding that $B_{\alpha,\lambda}$ decreases as λ increases, this property ensures that $B_{\alpha,\lambda}$ converges in probability to 0 as λ tends to infinity.

• As $(Y_x^{\alpha,0})_{x\in\mathbb{R}}$ oscillates when x tends to $-\infty$ (see [19] Corollary 2 on page 190), then

$$I_0^{\alpha,\lambda} \xrightarrow{\lambda\to 0} -\infty,$$

which ensures that $B_{\alpha,\lambda}$ converges in probability to 1 as λ tends to 0. □

Proof of Corollary 4.2.2. The first result is a direct consequence of Theorem 4.2.1.

If $o(1-\mathrm{m}t) = f_\alpha(x) \quad (x \to \infty)$, then for every $\lambda > 0$ and x large enough, $t \le (1-\lambda f_\alpha(x))/\mathrm{m}$ and

$$B_1(x,t)/x \le B_1(x, \frac{1-\lambda f_\alpha(x)}{\mathrm{m}})/x.$$

The right hand side converges weakly to $B_{\alpha,\lambda}$ as x tends to infinity. Letting λ tend to infinity, the lemma above entails that $B_1(x,t)/x \stackrel{x\to\infty}{\longrightarrow} 0$ in $\mathbb{P}$.

Similarly if $1-\mathrm{m}t = o(f_\alpha(x)) \quad (x \to \infty)$, then for every $\lambda > 0$ and x large enough,

$$B_1(x,t)/x \ge B_1(x, \frac{1-\lambda f_\alpha(x)}{\mathrm{m}})/x.$$

Letting λ tend to 0, Lemma 4.3.3 entails that $B_1(x,t)/x \stackrel{x\to\infty}{\longrightarrow} 1$ in $\mathbb{P}$. □

Chapter 5

Evolution of a typical data block

In this chapter, we focus on the dynamics of the covering and we shall study the block of data straddling a typical point, say 0 for simplicity, which is denoted by $\mathbf{B}_0$. Thus $\mathbf{B}_0(t)$ is the block of data of the hardware containing 0 at time t. We will show that its end-points and its length are pure jump Markov processes. Specifically, if a file arrives at time t at the left of $\mathbf{B}_0(t-)$ and cannot be stored entirely at its left, it yields a jump of the left end-point of $\mathbf{B}_0$. The data of this file which cannot be stored at the left of $\mathbf{B}_0(t-)$ are called *remaining data*. These remaining data yield a jump of the right end-point of $\mathbf{B}_0$ (see Figure 4). We shall prove that these events happen at instants which accumulate at 1/m and induce a random partition of the time interval $[0, 1/\mathrm{m}]$ with the Poisson-Dirichlet distribution (Theorem 5.2.1) and that the jumps of the end-points at these instants form a PPP on $[0, 1/\mathrm{m}] \times \mathbb{R}_+ \times \mathbb{R}_+$ (Proposition 5.4.1). Moreover the successive quantities of remaining data form an iid sequence (Corollary 5.3.2).
If a file arrives on $\mathbf{B}_0$, it yields a jump of the right end-point only (see Figure 5). The other files do not induce immediately a jump of $\mathbf{B}_0$ and we get the evolution of $(\mathbf{B}_0(t))_{t\geq 0}$ (Theorem 5.4.2). Finally, we prove that the process describing the length of $(\mathbf{B}_0(t))_{t\geq 0}$ is a branching process with immigration (Corollary 5.5.2).

Figure 4. Jumps of the end-points of $\mathbf{B}_0$ ($\Delta g(t)$ and $\Delta d(t)$) and remaining data induced by the arrival of a file at time t at the left of $\mathbf{B}_0(t^-)$.

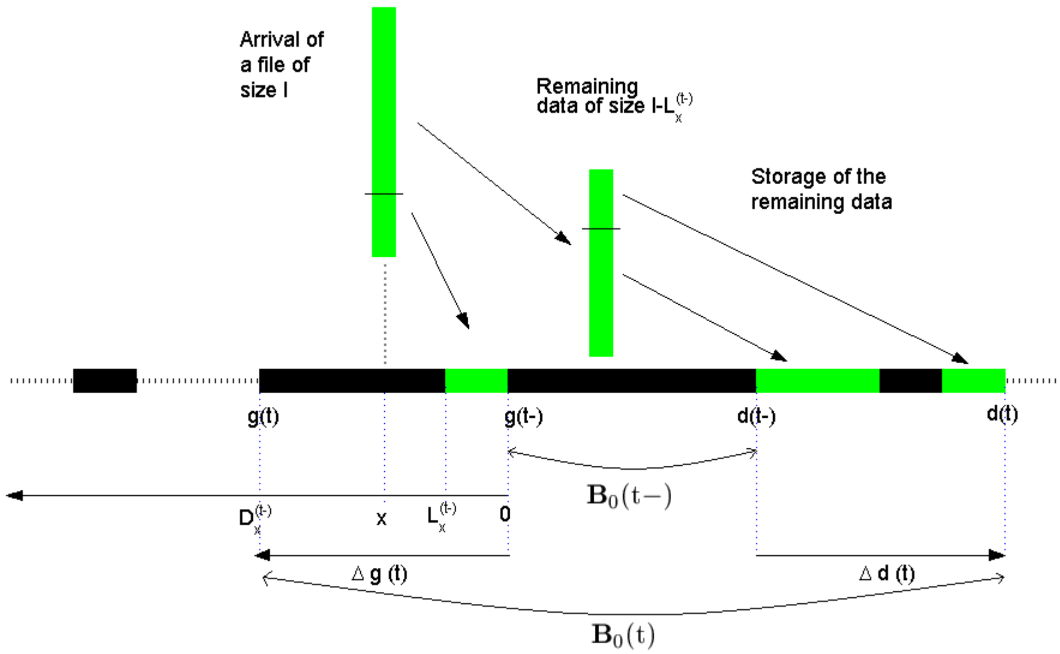
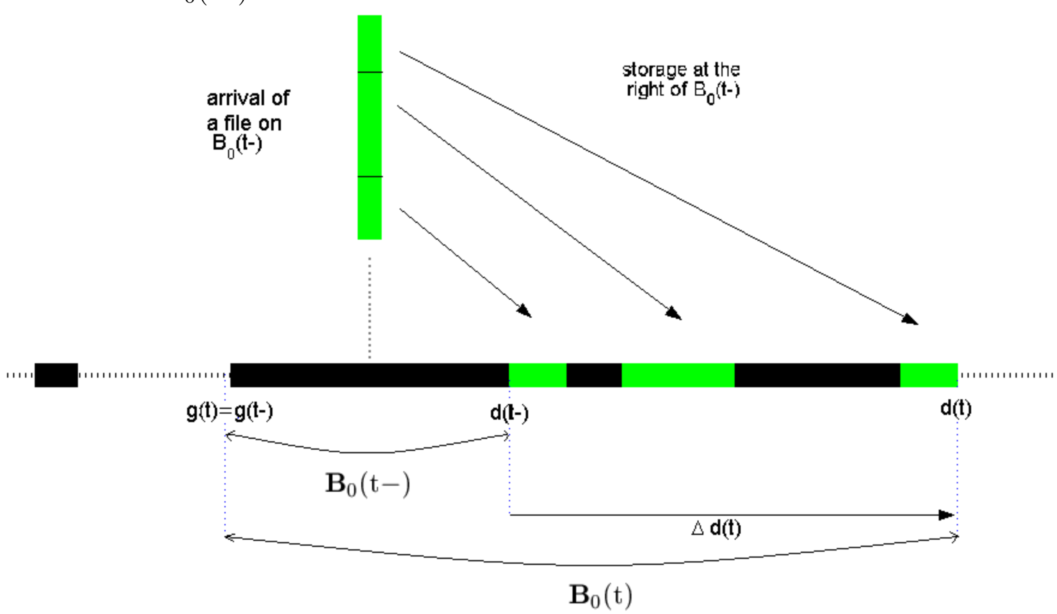


Figure 5. Jump of the right end-point of $\mathbf{B}_0$ $(\Delta d(t))$ induced by the arrival of a file at time t on $\mathbf{B}_0(t^-)$.



We use the same Notations as in the previous chapter and recall that $(Y_x^{(t)})_{x\geq 0}$

$$Y_0^{(t)} := 0 \quad ; \quad Y_b^{(t)} - Y_a^{(t)} = \sum_{\substack{t_i \leq t \\ x_i \in]a,b]}} l_i \quad - \quad (b-a) \quad \text{for } a < b, \tag{5.1}$$

is a Levy process with drift -1 and Lévy measure $t\nu(\mathrm{d}x)$.

In the previous chapter, we proved that the time when the hardware becomes full is equal to 1/m, that is a.s $\mathcal{C}(t) = \mathbb{R}$ iff $t \geq 1/\mathrm{m}$. Thus we already know that $\mathbf{B}_0(0) = \varnothing$ and $\mathbf{B}_0(1/\mathrm{m}) = \mathbb{R}$ and we shall study $(\mathbf{B}_0(t))_{t\in[0,1/\mathrm{m}]}$. In that view, we recall that $g(t)$ (resp. $d(t)$, resp. $l(t)$) is the left end-point (resp. the right end-point, resp. the length) of the data block containing 0 :

$$d(t) := d_0(\mathcal{R}(t)), \quad g(t) := -d_0(-\mathcal{R}(t)), \quad \mathbf{B}_0(t) := [g(t), d(t)[, \quad l(t) := d(t) - g(t).$$

We will also need the free space at the right of $\mathbf{B}_0(t)$ denoted by $\overrightarrow{\mathcal{R}(t)}$ and at the left of $\mathbf{B}_0(t)$, turned over, closed at the left and open at the right, denoted by $\overleftarrow{\mathcal{R}(t)}$, which were introduced in Section 3.3.1.

First, we prove some properties of absence of memory (Section 5.1) : the evolution of $\mathbf{B}_0$ after time t depends from the past of this block only through $l(t)$ (Markov property). Then we focus on the left end-point : it is an additive process and we give its Lévy measure. As a consequence, we get the distribution of the instants at which the left end-point jumps (Section 5.2) : these instants form a stick breaking sequence which does not depend on ν. We then derive the distribution of the remaining data (Section 5.3), which completes the description of the process of storage at the left end-point. By taking also into account the data fallen on $\mathbf{B}_0$, we get then the evolution of $(g(t), d(t))$ (Section 5.4). The latter characterizes the evolution of the right end-point and the length (Section 5.5).

5.1 Markov property of $\mathbf{B}_0$

We have already proved that $\mathcal{R}(t)$ enjoys a 'spatial' regeneration property (see Proposition 3.3.2). To study the evolution of $\mathbf{B}_0$, we need 'time' regeneration property. Here we prove that the evolution of the block containing 0 up to time t is independent of the covering outside $[g(t), d(t)]$ up to time t. In Section 5.4, this property will ensure that the evolution of $\mathbf{B}_0$ after time t depends from the past of this block only through $l(t)$ (Markov property).

Proposition 5.1.1. *For every $t \in [0, 1/\mathrm{m}[$, the following three processes with values in the space of subsets of $\mathbb{R}$*

. $\quad (g(t) - \mathcal{R}(s)) \cap [0, \infty[, \quad 0 \leq s \leq t,$

. $\quad (\mathcal{R}(s) - d(t)) \cap [0, \infty[, \quad 0 \leq s \leq t,$

. $\quad \mathcal{R}(s) \cap [g(t), d(t)], \quad 0 \leq s \leq t,$

are independent.

Remark 8. Actually, we have the following regeneration property : $\forall t \in [0, 1/\mathrm{m}[$, $\forall x \in \mathbb{R}$, $\big((\mathcal{R}(s) - d_x(\mathcal{R}(t))) \cap [0, \infty[: s \in [0, t]\big)$ is independent of $\big((\mathcal{R}(s) -$

$d_x(\mathcal{R}(t)))\cap]-\infty,0]: s\in[0,t]\big)$ and is distributed as $\big((\mathcal{R}(s)-d_0(\mathcal{R}(t)))\cap[0,\infty[: s\in[0,t]\big)$.

This result is a direct consequence of the following lemma where we consider the point processes of files until time t at the left of/at the right of/inside $[g,d]$:

$$P_g(t) := \{(t_i, g-x_i, l_i) : t_i \le t,\ x_i < g\}, \quad P^d(t) := \{(t_i, x_i-d, l_i) : t_i \le t,\ d < x_i\},$$

$$P_g^d(t) := \{(t_i, x_i, l_i) : t_i \le t,\ g \le x_i \le d\}.$$

Lemma 5.1.2. *For every $t \in [0, 1/\mathrm{m}[$, the point processes $P_{g(t)}(t)$, $P_{g(t)}^{d(t)}(t)$ and $P_{d(t)}(t)$ are independent.*

Proof. First we prove a weaker result, where times $(t_i)_{i\in\mathbb{N}}$ are not taken into account. Denote by $(\widetilde{Y}_x^{(t)})_{x\ge0}$ the càdlàg version of $(Y_{-x}^{(t)})_{x\ge0}$. This is a spectrally negative Lévy process with bounded variation, which drifts to ∞. Note that,

$$\begin{aligned} g(t) &= g_0(\mathcal{R}(t)) = \sup\{x \le 0:\ Y_x^{(t)} = I_x^{(t)}\} \\ &= \sup\{x \le 0:\ Y_{x^-}^{(t)} = I_0^{(t)}\} = -\inf\{x \ge 0:\ \widetilde{Y}_x^{(t)} = \inf\{\widetilde{Y}_z^{(t)} : z \ge 0\}\}. \end{aligned}$$

Then $(\widetilde{Y}_{-g(t)+x}^{(t)} - \widetilde{Y}_{-g(t)}^{(t)})_{x\ge0}$ is independent of $(\widetilde{Y}_x^{(t)})_{0\le x\le -g(t)}$ (decomposition of a Lévy process at its infimum [74]). Considering the locations and sizes of the jumps of these two processes yields

$$\{(g(t)-x_i, l_i) : t_i \le t,\ x_i < g(t)\} \quad \text{is independent of} \quad \{(x_i, l_i) : t_i \le t,\ g(t) \le x_i \le 0\}.$$

Adding that $\{(x_i, l_i) : t_i \le t,\ x_i > 0\}$ is independent of $\{(x_i, l_i) : t_i \le t,\ x_i \le 0\}$ and $g(t)$ is $\{(x_i, l_i) : t_i \le t,\ x_i \le 0\}$ measurable, we get

$$\{(g(t)-x_i, l_i) : t_i \le t,\ x_i < g(t)\} \quad \text{is independent of} \quad \{(x_i, l_i) : t_i \le t,\ x_i \ge g(t)\}.$$

We now extend the preceding by incorporating the times $(t_i)_{i\in\mathbb{N}}$. In this direction, we recall that if $(\widetilde{x}_i, \widetilde{l}_i)_{i\in\mathbb{N}}$ is a PPP on $\mathbb{R}\times\mathbb{R}_+$ with intensity $t\mathrm{d}x \otimes \nu(\mathrm{d}l)$ and $(\widetilde{t}_i)_{i\in\mathbb{N}}$ is an iid sequence distributed uniformly on $[0,t]$, then $\{(\widetilde{t}_i, \widetilde{x}_i, \widetilde{l}_i) : i \in \mathbb{N}\}$ is distributed as $\{(t_i, x_i, l_i) : i \in \mathbb{N}, t_i \le t\}$. Adding that $g(t)$ is $\{(x_i, l_i) : i \in \mathbb{N}, t_i \le t\}$ measurable, we get

$$\{(t_i, g(t)-x_i, l_i) : t_i \le t,\ x_i < g(t)\} \quad \text{is independent of} \quad \{(t_i, x_i, l_i) : t_i \le t,\ x_i \ge g(t)\}.$$

This ensures that $P_{g(t)}(t)$ is independent of $(P_{g(t)}^{d(t)}(t), P^{d(t)}(t))$.

One can prove similarly that $P^{d(t)}(t)$ is independent of $(P_{g(t)}(t), P_{g(t)}^{d(t)}(t))$ using that $(Y_{d(t)+x}^{(t)} - Y_{d(t)}^{(t)})_{x\ge0}$ is independent of $(Y_x^{(t)})_{x\le d(t)}$ or Lemma 2 in [13]. □

This guarantees the absence of memory at the left of $\mathbf{B}_0(t)$. First we have :

Corollary 5.1.3. $(g(t))_{t\in[0,1/\mathrm{m}]}$ *has decreasing càdlàg paths with independent increments.*

Proof. Let $0 \le t < t+s \le 1/\mathrm{m}$. The increment $g(t+s)-g(t)$ just depends on $\overleftarrow{\mathcal{R}(t)}$ and the point process of files which arrive after time t at the left of $\mathbf{B}_0(t)$ $\{(t_i, x_i - g(t), l_i) : t_i > t, x_i < g(t)\}$. By the Poissonian property, these two quantities are independent and $(g(u) : u \in [0,t])$ is independent of this point process of files. Moreover $(g(u) : u \in [0,t])$ is also independent of $(g(t) - \mathcal{R}(t)) \cap [0,\infty[$ by Proposition 5.1.1. So $(g(u) : u \in [0,t])$ is independent of $g(t+s) - g(t)$. □

This explains the observation made in Section 3.3 that the distribution of $g(t)$ is infinitively divisible (see [36] on page 174 or [86] on page 47 for details).

5.2 Evolution of the left end-point

Now we describe the process $(g(t))_{t\in[0,1/\mathrm{m}[}$. We know that its increments are independent and (3.13) specifies its marginals. We shall determine its Lévy measure and prove that its mass is finite (see [86] for terminology). This means that the instants when a file arrives at the left of $\mathbf{B}_0$ and joins this data block during its storage do not accumulate before time $1/\mathrm{m}$, even if $\bar{\nu}(0) = \infty$ (files arrive densely near the data block). Proposition 3 in [13] ensures that the first time T_1 when 0 is covered, which is also the first jump time of $(g(t))_{t\in[0,1/\mathrm{m}]}$, is uniformly distributed on $[0, 1/\mathrm{m}]$. Actually the second jump time is uniformly distributed in $[T_1, 1/\mathrm{m}]$ and so on ... More precisely, we have :

Theorem 5.2.1. *The jump times of* $(g(t))_{t\in[0,1/\mathrm{m}]}$ *are given by an increasing sequence* $(T_i)_{i\in\mathbb{N}}$ *which accumulate at* $1/\mathrm{m}$. *More precisely, using the convention* $T_0 = 0$, *it holds that for every* $i \ge 1$, *conditionally on* $T_{i-1} = t$, T_i *is independent of* $(T_j)_{0\le j\le i-1}$ *and is uniformly distributed on* $[t, 1/\mathrm{m}]$.

Then, denoting by $-G_i$ *the jump of* $(g(t))_{t\in[0,1/\mathrm{m}]}$ *at time* T_i *for every* $i \in \mathbb{N}$, *we have*

$$g(t) := -\sum_{T_i \le t} G_i$$

where $\{(T_i, G_i) : i \in \mathbb{N}\}$ *is a PPP on* $[0, 1/\mathrm{m}[\times\mathbb{R}^+$ *with intensity*

$$\mathrm{d}t\mathrm{d}x \int_0^\infty \mathbb{P}(Y_x^{(t)} \in -\mathrm{d}l)\bar{\nu}(l).$$

In other words, $(g(t))_{t\in[0,1/\mathrm{m}]}$ is an additive process and its generating triplet is

$$\left(0,\int_0^t \mathrm{d}s \int_0^\infty \mathbb{P}(Y_x^{(s)} \in -\mathrm{d}l)\bar{\nu}(l), 0\right).$$

In particular, the interarrival times of $\{T_i : i \in \mathbb{N}\}$ form a 'continuous uniform stick breaking sequence' (see the residual allocation model in [80] on pages 63-64) : the distribution of $\big((T_{i+1}-T_i)/\mathrm{m}\big)_{i\in\mathbb{N}}$ is the Griffiths-Engen-McCloskey distribution with parameter $(0,1)$ (i.e. rearranging these increments in the decreasing order yields the Poisson-Dirichlet distribution of parameter $(0,1)$).

Further, for every $i \in \mathbb{N}$, conditionally on $T_i = t$, the law of G_i is given by

$$\mathbb{P}(G_i \in \mathrm{d}x) = \mathrm{d}x\frac{1-\mathrm{m}t}{\mathrm{m}}\int_0^\infty \mathbb{P}(Y_x^{(t)} \in -\mathrm{d}l)\bar{\nu}(l), \tag{5.2}$$

and as a consequence,

$$\mathbb{E}(G_i) = \big(\frac{1}{(1-\mathrm{m}t)^2} + \frac{1}{2}\frac{\mathrm{m}}{1-\mathrm{m}t}\big)\int_0^\infty l^2\nu(\mathrm{d}l).$$

Example 3. For the basic example ($\nu = \delta_1$), conditionally on $T_i = t$, we have,

$$\mathbb{P}(G_i \in \mathrm{d}x) = (1-t)e^{-tx}\frac{(tx)^{[x]}}{[x]!}\mathrm{d}x,$$

writing $[x] = \sup\{n \in \mathbb{N} : n \leq x\}$ and using (3.11).

For the proof, we need the following identity

Lemma 5.2.2. *Let $(S_t)_{t\geq 0}$ be a subordinator with no drift and Lévy tail $\bar{\mu}$. Then for all $(t,x) \in \mathbb{R}_+^2$, we have*

$$\mathbb{P}(S_t > x) = \int_0^t \mathrm{d}s \int_0^x \mathbb{P}(S_s \in \mathrm{d}b)\bar{\mu}(x-b).$$

Proof. As S has no drift, we have for all $t > 0$ and $x > 0$,

$$S_t > x \quad \Leftrightarrow \quad \exists!\ s \in]0,t]\ :\ S_{s^-} \leq x,\ \Delta S_s > x - S_{s^-} \qquad \text{a.s.}$$

We get then, using also the compensation formula (see [19] on page 7),

$$\mathbb{P}(S_t > x) = \mathbb{E}(\sum_{0<s\leq t} \mathbb{1}_{\{S_{s^-}\leq x\}}\mathbb{1}_{\{\Delta S_s > x - S_{s^-}\}}) = \mathbb{E}(\int_0^t \mathrm{d}s \mathbb{1}_{\{S_s \leq x\}}\bar{\mu}(x-S_s))$$

which completes the proof. One can also give an analytic proof by computing the Laplace transform of the right hand side for $q > 0$. Using Fubini and denoting by ϕ the Laplace transform of $(S_t)_{t\geq 0}$:

$$\begin{aligned}
& \int_0^\infty \mathrm{d}x e^{-qx} \int_0^t \mathrm{d}s \int_0^x \mathbb{P}(S_s \in \mathrm{d}b)\bar{\mu}(x-b) \\
= & \int_0^t \mathrm{d}s \int_0^\infty \mu(\mathrm{d}y) \int_0^\infty \mathbb{P}(S_s \in \mathrm{d}b) \frac{e^{-qb} - e^{-q(b+y)}}{q} \\
= & \int_0^t \mathrm{d}s e^{-\phi(q)s} \int_0^\infty \mu(\mathrm{d}y) \frac{1-e^{-qy}}{q} \\
= & \frac{1-e^{-\phi(q)t}}{\phi(q)} \times \frac{\phi(q)}{q} = \int_0^\infty \mathrm{d}x e^{-qx} \mathbb{P}(S_t > x)
\end{aligned}$$

which proves the lemma. □

We are now able to establish Theorem 5.2.1.

Proof. We know from Corollary 5.1.3 that $(g(t))_{t\in[0,1/\mathrm{m}]}$ is an additive process. Moreover for every $x \geq 0$, $(Y_x^{(t)} + x)_{t\geq 0}$ is a subordinator with no drift and Lévy measure $x\nu$ (see (5.1)). So Lemma 5.2.2 ensures that

$$\begin{aligned}
\mathbb{P}(Y_x^{(t)} > 0) &= \mathbb{P}(Y_x^{(t)} + x > x) \\
&= \int_0^t \mathrm{d}s \int_0^x \mathbb{P}(Y_x^{(s)} + x \in \mathrm{d}b) x\bar{\nu}(x-b) \\
&= \int_0^t \mathrm{d}s \int_0^\infty \mathbb{P}(Y_x^{(s)} \in -\mathrm{d}l) x\bar{\nu}(l).
\end{aligned}$$

Using (3.13), we get

$$\mathbb{E}\big(\exp\big(\lambda g(t)\big)\big) = \exp\bigg(\int_0^\infty \mathrm{d}x(e^{-\lambda x} - 1) \int_0^t \mathrm{d}s \int_0^\infty \mathbb{P}(Y_x^{(s)} \in -\mathrm{d}l)\bar{\nu}(l)\bigg).$$

So $(g(t))_{t\in[0,1/\mathrm{m}]}$ is an additive process with generating triplet

$$\bigg(0, \int_0^t \mathrm{d}s \int_0^\infty \mathbb{P}(Y_x^{(s)} \in -\mathrm{d}l)\bar{\nu}(l), 0\bigg)$$

using Definition 8.2 and Theorem 9.8 in [86]. This characterizes the distribution of $(g(t))_{t\in[0,1/\mathrm{m}]}$ (by Theorem 9.8 in [86]) and proves that $\{(T_i, G_i) : i \in \mathbb{N}\}$ is a PPP on $[0, 1/\mathrm{m}[\times\mathbb{R}^+$ with intensity $\mathrm{d}t\mathrm{d}x \int_0^\infty \mathbb{P}(Y_x^{(t)} \in -\mathrm{d}l)\bar{\nu}(l)$. One can also compute the distribution of $g(t+s) - g(t)$ using the independence of increments and (3.13) : this proves that $g(.)$ is the sum of jumps given by a PPP.

By projection, $\{T_i : i \in \mathbb{N}\}$ is a PPP on $[0, 1/\mathrm{m}[$ with intensity $\mathrm{m}(1-\mathrm{m}t)^{-1}\mathrm{d}t$. Indeed, for every $t \in [0, 1/\mathrm{m}[$,

$$\begin{aligned}
\int_0^\infty \mathrm{d}x \int_0^\infty \mathbb{P}(Y_x^{(t)} \in -\mathrm{d}l)\bar{\nu}(l) &= \int_0^\infty \mathbb{P}(\overrightarrow{\tau}_l^{(t)} \in \mathrm{d}x) \int_0^\infty \mathrm{d}l \frac{x}{l}\bar{\nu}(l) \quad \text{using (3.9)} \\
&= \int_0^\infty \mathrm{d}l \frac{\mathbb{E}(\overrightarrow{\tau}_l^{(t)})\bar{\nu}(l)}{l} \\
&= \mathbb{E}(\overrightarrow{\tau}_1^{(t)}) \int_0^\infty \bar{\nu}(l)\mathrm{d}l \\
&= \frac{\mathrm{m}}{1-\mathrm{m}t} \quad \text{using (2.15).}
\end{aligned}$$

Thus, writing $N_t^{t'} := \mathrm{card}\{i \in \mathbb{N} : T_i \in]t, t']\}$, we have $N_0^t < \infty$ a.s. for every $t \in [0, 1/\mathrm{m}[$. We we can then sort the times T_i and we have

$$\mathbb{P}(T_{i+1} > t' \mid T_i = t) = \mathbb{P}(N_t^{t'} = 0) = \exp\big(- \int_t^{t'} \mathrm{d}s \frac{\mathrm{m}}{1-\mathrm{m}s}\big) = \frac{1-\mathrm{m}t'}{1-\mathrm{m}t},$$

meaning that T_{i+1} is uniformly distributed in $[T_i, 1/\mathrm{m}]$. The independence is a consequence of the Poissonian property of $\{T_i : i \in \mathbb{N}\}$ and we get the theorem.

Finally, this proves (5.2) and for every $i \in \mathbb{N}$, conditionally on $T_i = t$, we get

$$\begin{aligned}
\mathbb{E}(G_i) &= \frac{1-\mathrm{m}t}{\mathrm{m}} \int_0^\infty \mathrm{d}l \frac{\mathbb{E}([\overrightarrow{\tau}_l^{(t)}]^2)\bar{\nu}(l)}{l} \quad \text{using again (3.9)} \\
&= \frac{1-\mathrm{m}t}{\mathrm{m}} \int_0^\infty \mathrm{d}l \bar{\nu}(l) \big(l \big(\frac{\mathrm{m}}{1-\mathrm{m}t}\big)^2 + \frac{\int_0^\infty l^2 \nu(\mathrm{d}l)}{(1-\mathrm{m}t)^3} \big)
\end{aligned}$$

since $[\kappa^{(t)}]'(0)$ is given by (2.15) and $[\kappa^{(t)}]''(0)$ is given by Proposition 4 in [13]. $\square$

5.3 The process of remaining data

We still consider the files which arrive at the left of $\mathbf{B}_0$, the block containing 0, and cannot be entirely stored at the left of this block (see Figure 4). Such events occur at the jump times of $(g(t))_{t \in [0,1/\mathrm{m}]}$, that is at time T_i. We focus here on the portions of these files which cannot be stored at the left of $\mathbf{B}_0$ and are shifted to the right of $\mathbf{B}_0(T_i-)$ to find a free space. They are called remaining data and denoted by R_i. Thus R_i is the quantity of data which arrives at the left of $\mathbf{B}_0$ at time T_i and is stored at the right of $\mathbf{B}_0$. Then it is also the quantity of data over $g(T_{i-1}-)$ at time T_i (see Section 3.2 for details) and it is given by

$$\forall i \geq 1, \qquad R_i := Y^{(T_i)}_{g(T_{i-1}-)} - I^{(T_i)}_{g(T_{i-1}-)}.$$

We aim at determining the distribution of $\{(T_i, G_i, R_i) \ : \ i \in \mathbb{N}\}$ which is the key to the characterization of the jumps of $(g(t), d(t))_{t\in[0,1/\mathrm{m}]}$. In that view, we need to describe the arrival of files which induce the jumps (G_i, R_i). So we consider the half hardware at the left of $g(t)$, which we turn over, so that it is now identified with $\mathbb{R}^+$ and its free space is given by $\overleftarrow{\mathcal{R}(t)}$ (see Section 3.3). The size of free space and the first free plots of this half hardware are given by the processes $(L^{(t)}_x)_{x\geq 0}$ and $(D^{(t)}_x)_{x\geq 0}$ defined by

$$\forall t \in [0, 1/\mathrm{m}[, \ \forall x \geq 0, \qquad L^{(t)}_x = | \ \overleftarrow{\mathcal{R}(t)} \cap [0, x] \ |, \qquad D^{(t)}_x = \inf\{y > x : y \in \overleftarrow{\mathcal{R}(t)}\}.$$

When at time t, a file of length l arrives at location $-x + g(t-)$ on the hardware (i.e. at location x on the half hardware), it yields a jump of $g(.)$ if the free space $L^{(t-)}_x$ between $-x+g(t-)$ and $g(t-)$ is less than l. Then the quantity of remaining data is $l - L^{(t-)}_x$ and the jump of the left end-point is $D^{(t-)}_x$ (see Figure 4). So we naturally introduce the measure $\rho^{(t)}$ on $\mathbb{R}^2_+$ defined by

$$\rho^{(t)}(\mathrm{d}y\mathrm{d}z) := \int_0^\infty \mathrm{d}x \int_0^\infty \nu(\mathrm{d}l)\mathbb{P}(D^{(t)}_x \in \mathrm{d}y, l - L^{(t)}_x \in \mathrm{d}z).$$

In forthcoming Lemma 5.3.3, we give a useful alternative expression of $\rho^{(t)}$. This measure gives the intensity of the point process $\{(T_i, G_i, R_i) : i \in \mathbb{N}\}$, as stated by the following result.

Theorem 5.3.1. *$\{(T_i, G_i, R_i) : i \in \mathbb{N}\}$ is a PPP on $[0, 1/\mathrm{m}[\times\mathbb{R}^2_+$ with intensity $\mathrm{d}t\rho^{(t)}(\mathrm{d}y\mathrm{d}z)$.*

A remarkable consequence is that $(R_i)_{i\in\mathbb{N}}$ is an iid sequence : whereas the rate at which jumps occur increases as time gets closer to $1/\mathrm{m}$, the quantity of remaining data keeps the same distribution.

Corollary 5.3.2. *$\{(T_i, R_i) \ : \ i \in \mathbb{N}\}$ is a PPP on $[0, 1/\mathrm{m}[\times\mathbb{R}^+$ with intensity $\mathrm{d}t\mathrm{d}z\frac{\bar{\nu}(z)}{1-\mathrm{m}t}$.*
In other words, $(R_i)_{i\in\mathbb{N}}$ is iid, independent of $(T_i)_{i\in\mathbb{N}}$ and its distribution is given by :

$$\mathbb{P}(R_i \in \mathrm{d}z) = \mathrm{m}^{-1}\bar{\nu}(z)\mathrm{d}z, \quad z \geq 0.$$

Example 4. Using the expression of $\rho^{(t)}$ given by Lemma 5.3.3 below, the expressions (23) and (24) in [13] yield an expression of $\rho^{(t)}$ for the basic example and the gamma distribution which is quite heavy and not mentioned here. Nonetheless the quantity of remaining data can be often calculated explicitly. For the basic example ($\nu = \delta_1$), the remaining data are uniform random variables on $[0, 1]$. For the exponential distribution ($\nu(\mathrm{d}l) = \mathbb{1}_{\{l\geq 0\}}e^{-l}\mathrm{d}l$), the remaining data are also exponentially distributed.

The proofs of these results are organized as follows.
First, in Lemma 5.3.3, we give a more explicit expression of $\rho^{(t)}$ which will be useful for the proofs and will enable us to derive Corollary 5.3.2 from Theorem 5.3.1.
Second, we prove that $\rho^{(t)}$ gives the intensity of the point process $\{(T_i, G_i, R_i) : i \in \mathbb{N}\}$ (Lemma 5.3.4). That is for every $t \in [0, 1/\mathrm{m}[$ and $A =]a_1, b_1] \times]a_2, b_2] \subset \mathbb{R}^2_+$, we have :

$$\lim_{h\to 0} \frac{\mathbb{P}(\exists i \in \mathbb{N} : T_i \in]t, t+h],\ (G_i, R_i) \in A)}{h} = \rho^{(t)}(A).$$

The lower bound appears naturally by considering the arrival of one single file independently of the past which induces a jump of the left end-point, as described at the beginning of this section (see also Figure 4). However, in the case $\bar{\nu}(0) = \infty$, some jumps of the left end-point could be due to the successive arrival of many files during a short time interval $]t, t+h]$. Thanks to Theorem 5.2.1, we already know the rate at which jumps occur (i.e. the total intensity). This will give us the upper bound.
Finally, we prove that the point process $\{(T_i, G_i, R_i) : i \in \mathbb{N}\}$ enjoys a memoryless property (Lemma 5.3.5), which is a direct consequence of results of Section 5.1. We get then the complete description of this point process, which enables us to prove Theorem 5.3.1. Corollary 5.3.2 follows by integrating $\rho^{(t)}$ with respect to the first coordinate.

Recall the notation in Proposition 3.3.3 and (3.7).

Lemma 5.3.3. *For every* $t \in [0, 1/\mathrm{m}[$, *the measure* $\rho^{(t)}(\mathrm{d}y\mathrm{d}z)$ *can also be expressed as*

$$\begin{aligned}
&\mathrm{d}z \int_z^\infty \nu(\mathrm{d}l)\left(\mathbb{P}(\overleftarrow{\tau}^{(t)}_{l-z} \in \mathrm{d}y) + \int_0^y \mathbb{P}(\overleftarrow{\tau}^{(t)}_{l-z} \in \mathrm{d}x)(y-x)\Pi^{(t)}(\mathrm{d}y - x)\right) \\
=& \int_z^\infty \nu(\mathrm{d}l)(l-z)\left(y^{-1}\mathrm{d}y\mathbb{P}(Y^{(t)}_y + l \in \mathrm{d}z) + \int_0^y \mathbb{P}(Y^{(t)}_x + l \in \mathrm{d}z)(yx^{-1} - 1)\Pi^{(t)}(\mathrm{d}y - x)\right)
\end{aligned}$$

Proof. By Lemma 1.11 in Chapter 1 of [21] applied to $(\overleftarrow{\tau}^{(t)}_x)_{x\geq 0}$, we have for all $a, b \geq 0$ and $q > 0$ (t is fixed and omitted in the notation),

$$\int_0^\infty \mathrm{d}x e^{-qx}\mathbb{E}(\exp(-bL_x - aD_x)) = \frac{\kappa(a+q) - \kappa(a)}{q(\kappa(a+q) + b)}.$$

Letting $q \to 0$, we get

$$\int_0^\infty \mathrm{d}x\mathbb{E}(\exp(-bL_x - aD_x)) = \frac{\kappa'(a)}{\kappa(a)+b} = \int_0^\infty \mathrm{d}z e^{-bz}\kappa'(a)e^{-\kappa(a)z}.$$

From $\kappa'(a) = \int_0^\infty e^{-ay}(\delta_0(\mathrm{d}y) + y\Pi(\mathrm{d}y))$ and $e^{-\kappa(a)z} = \int_0^\infty e^{-ay}\mathbb{P}(\overleftarrow{\tau}_z \in \mathrm{d}y)$, we deduce

$$\int_0^\infty \mathrm{d}x\mathbb{E}(\exp(-bL_x - aD_x)) = \int_0^\infty \mathrm{d}z \int_0^\infty \gamma_z(\mathrm{d}y)e^{-bz-ay}, \tag{5.3}$$

where γ_z is the convolution of $\delta_0(\mathrm{d}y) + y\Pi(\mathrm{d}y)$ and $\mathbb{P}(\overleftarrow{\tau}_z \in \mathrm{d}y)$. Thus,

$$\begin{aligned}\gamma_z(\mathrm{d}y) &= \int_0^y \mathbb{P}(\overleftarrow{\tau}_z \in \mathrm{d}x)(\delta_0(\mathrm{d}y - x) + (y-x)\Pi(\mathrm{d}y - x)) \\ &= \mathbb{P}(\overleftarrow{\tau}_z \in \mathrm{d}y) + \int_0^y \mathbb{P}(\overleftarrow{\tau}_z \in \mathrm{d}x)(y-x)\Pi(\mathrm{d}y-x).\end{aligned}$$

And the identification of Laplace transforms in (5.3) entails that

$$\int_0^\infty \mathrm{d}x\mathbb{P}(L_x \in \mathrm{d}z, D_x \in \mathrm{d}y) = \mathrm{d}z\big(\mathbb{P}(\overleftarrow{\tau}_z \in \mathrm{d}y) + \int_0^y \mathbb{P}(\overleftarrow{\tau}_z \in \mathrm{d}x)(y-x)\Pi(\mathrm{d}y-x)\big), \tag{5.4}$$

which proves the first identity of the lemma integrating with respect to l. Using (3.9) gives the second one. □

Remark 9. A recent work of Winkel (Theorem 1 in [92]) enables us to calculate differently the law of $\mathbb{P}(L_x \in \mathrm{d}z, D_x \in \mathrm{d}y)$ (L_x corresponds to T_x in [92] and D_x to $X(T_{x-}) + \Delta_x$) :

$$\int_0^\infty \mathrm{d}x\mathbb{P}(L_x \in \mathrm{d}z, D_x \in \mathrm{d}y) = \mathrm{d}y\mathbb{P}(H_y \in \mathrm{d}z) + \mathrm{d}z\int_0^\infty \mathbb{P}(\overleftarrow{\tau}_x \in \mathrm{d}x)(y-x)\Pi(\mathrm{d}y-x),$$

where $H_x = \inf\{a \geq 0, \overleftarrow{\tau}_a = x\}$. Then observe that the measures on $\mathbb{R}_+^2$ $\mathrm{d}y\mathbb{P}(H_y \in \mathrm{d}z)$ and $\mathrm{d}z\mathbb{P}(\overleftarrow{\tau}_z \in \mathrm{d}y)$ coincide by computing their Laplace transform using (4) in [92]. This proves (5.4).

Second, for every Borel set B of $[0, 1/\mathrm{m}[\times\mathbb{R}_+^2$, we define $N_B := \mathrm{card}\{i \in \mathbb{N} : (T_i, G_i, R_i) \in B\}$ and we say that A is a rectangle of $D \subset \mathbb{R}^{\mathrm{d}}$ if A is a subset of D of the form

$$\{x = (x_1, x_2, .., x_{\mathrm{d}}), a_1 < x_1 \leq b_1, .., a_{\mathrm{d}} < x_{\mathrm{d}} \leq b_{\mathrm{d}}\}.$$

Then, we have

Lemma 5.3.4. *For all $t \in [0, 1/\mathrm{m}[$ and A rectangle of $\mathbb{R}_+^2$, we have :*

$$\lim_{h\to 0} \frac{\mathbb{P}(N_{]t,t+h]\times A} \geq 1)}{h} = \rho^{(t)}(A).$$

Proof. First we prove the lower bound. Second, we check that the convergence holds for $A = \mathbb{R}_+^2$.

• Let $\epsilon > 0$, $A =]a,b]\times]c,d]$ and work conditionally on $\overleftarrow{\mathcal{R}(t)}$. We consider a file labelled i which arrives at time $t_i \in]t, t+h]$ at location $x_i < g(t)$. We put $\widetilde{x_i} := g(t) - x_i \geq 0$ the arrival point on the half line at the left of $g(t)$ and require that

$$l_i - L_{\widetilde{x_i}}^{(t)} \in]c, d-\epsilon],\ D_{\widetilde{x_i}}^{(t)} \in]a, b-\epsilon],\ |L_{\widetilde{x_i}}^{(t_i-)} - L_{\widetilde{x_i}}^{(t)}| \leq \epsilon,\ |D_b^{(t_i-)} - D_b^{(t)}| \leq \epsilon.$$

Then file i verifies

$$l_i - L_{\widetilde{x_i}}^{(t_i-)} \in]c, d],\ D_{\widetilde{x_i}}^{(t_i-)} \in]a, b].$$

So this file induces a jump of the left end-point and $N_{]t,t+h]\times A} \geq 1$ (see the beginning of this section or Figure 4 for details) and we get the lower bound :

$$\begin{aligned}
&\mathbb{P}\big(N_{]t,t+h]\times A} \geq 1 \ \mid\ \overleftarrow{\mathcal{R}(t)}\big) \\
&\qquad \geq\ \mathbb{P}\big(\exists i \in \mathbb{N}\ :\ t_i \in]t, t+h],\ l_i - L_{\widetilde{x_i}}^{(t)} \in]c, d-\epsilon],\ D_{\widetilde{x_i}}^{(t)} \in]a, b-\epsilon], \\
&\qquad\qquad |L_{\widetilde{x_i}}^{(t_i-)} - L_{\widetilde{x_i}}^{(t)}| \leq \epsilon,\ |D_b^{(t_i-)} - D_b^{(t)}| \leq \epsilon \ \mid\ \overleftarrow{\mathcal{R}(t)}\big) \\
&\qquad \geq\ A_t(h).B_t(h) \qquad (5.5)
\end{aligned}$$

where

$$\begin{aligned}
A_t(h) &:= \mathbb{P}\big(\exists i \in \mathbb{N} :\ t_i \in]t, t+h], l_i - L_{\widetilde{x_i}}^{(t)} \in]c, d-\epsilon], D_{\widetilde{x_i}}^{(t)} \in]a, b-\epsilon] \ \mid\ \overleftarrow{\mathcal{R}(t)}\big), \\
B_t(h) &:= \mathbb{P}\big(\sup_{t' \in [t,t+h]} \{|L_b^{(t')} - L_b^{(t)}|\} \leq \epsilon,\ \sup_{t' \in [t,t+h]} \{|D_b^{(t')} - D_b^{(t)}|\} \leq \epsilon \ \mid\ \overleftarrow{\mathcal{R}(t)}\big).
\end{aligned}$$

1) By Theorem 5.2.1, $\mathbb{P}(N_t^{t+h} \neq 0) \xrightarrow{h\to 0} 0$ so a.s for h small enough, $g(t+h) = g(t)$. Then, using the Hausdorff metric on $\mathbb{R}_+$ (denoted by $\mathcal{H}(\mathbb{R}_+)$ in Chapter 2), we have

$$\overleftarrow{\mathcal{R}(t+h)} \xrightarrow{h\to 0} \overleftarrow{\mathcal{R}(t)} \quad \text{a.s.}$$

Then $B_t(h)$ converges a.s. to 1 as h tends to 0.

2) As $\{(t_i, \widetilde{x_i}, l_i) : i \in \mathbb{N}, t_i \in]t, t+h], x_i < g(t)\}$ is a PPP on $]t, t+h] \times \mathbb{R}_+^2$ with intensity $\mathrm{d}t \otimes \mathrm{d}x \otimes \nu(\mathrm{d}l)$ independent of $\overleftarrow{\mathcal{R}(t)}$,

$$A_t(h) = 1 - \exp\big(-h \int_0^\infty \mathrm{d}x \int_0^\infty \nu(\mathrm{d}l) \mathbb{1}_{\{l - L_x^{(t)} \in]c,d-\epsilon], D_x^{(t)} \in]a,b-\epsilon]\}}\big) \qquad \text{a.s.}$$

This term is a.s. equivalent when h tends to 0 to

$$h \int_0^\infty \mathrm{d}x \int_0^\infty \nu(\mathrm{d}l) \mathbb{1}_{\{l - L_x^{(t)} \in]c,d-\epsilon], D_x^{(t)} \in]a,b-\epsilon]\}}.$$

Then, letting $h \to 0$ in (5.5), 1) and 2) give

$$\liminf_{h\to 0} \frac{\mathbb{P}\big(N_{]t,t+h]\times A} \geq 1 \ \mid \ \overleftarrow{\mathcal{R}(t)}\big)}{h} \geq \int_0^\infty \mathrm{d}x \int_0^\infty \nu(\mathrm{d}l) \mathbb{1}_{\{l-L_x^{(t)}\in]c,d-\epsilon], D_x^{(t)}\in]a,b-\epsilon]\}} \quad \text{a.s.}$$

Integrating this inequality and using Fatou's lemma yield

$$\begin{aligned}\liminf_{h\to 0} \frac{\mathbb{P}\big(N_{]t,t+h]\times A} \geq 1\big)}{h} &\geq \mathbb{E}\Big(\int_0^\infty \mathrm{d}x \int_0^\infty \nu(\mathrm{d}l) \mathbb{1}_{\{l-L_x^{(t)}\in]c,d-\epsilon], D_x^{(t)}\in]a,b-\epsilon]\}}\Big) \\ &\geq \rho^{(t)}(]a,b-\epsilon]\times]c,d-\epsilon]).\end{aligned}$$

As $\rho^{(t)}(]a,b]\times\{d\}\cup\{b\}\times]c,d]) = 0$ (use the two equalities of Lemma 5.3.3), we get letting ϵ tend to 0 :

$$\liminf_{h\to 0} \frac{\mathbb{P}\big(N_{]t,t+h]\times A} \geq 1\big)}{h} \geq \rho^{(t)}(A).$$

- We derive the upper bound from Theorem 5.2.1. First,

$$\frac{\mathbb{P}\big(N_{]t,t+h]\times\mathbb{R}_+^2} \geq 1\big)}{h} = \frac{\mathbb{P}(\exists i \in \mathbb{N} \ : \ T_i \in]t,t+h])}{h} \overset{h\to 0}{\longrightarrow} \frac{\mathrm{m}}{1-\mathrm{m}t}.$$

and identity (5.7) below gives

$$\rho^{(t)}(\mathbb{R}_+^2) = \frac{\mathrm{m}}{1-\mathrm{m}t}.$$

So we just need to prove the following result : Let $(\mu_n)_{n\in\mathbb{N}}$ and μ be finite measures on $\mathbb{R}_+^2$ such that for every A rectangle of $\mathbb{R}_+^2$: $\liminf_{n\to\infty}\mu_n(A) \geq \mu(A)$ and $\lim_{n\to\infty}\mu_n(\mathbb{R}_+^2) = \mu(\mathbb{R}_+^2)$. Then for every A rectangle of $\mathbb{R}_+^2$, $\lim_{n\to\infty}\mu_n(A) = \mu(A)$.
In that view, suppose there exist a rectangle A, $\epsilon > 0$ and a sequence of integers k_n such that $\mu_{k_n}(A) \geq \mu(A)+\epsilon$. Choose B union of disjoint rectangles all disjoint from A such that $\mu(B\cup A) \geq \mu(\mathbb{R}_+^2) - \epsilon/2$. Then,

$$\liminf_{n\to\infty}\mu_{k_n}(\mathbb{R}_+^2) \geq \liminf_{n\to\infty}\mu_{k_n}(A\cup B) \geq \mu(A)+\epsilon+\mu(B) \geq \mu(\mathbb{R}_+^2)+\epsilon/2,$$

which is a contradiction with $\lim_{n\to\infty}\mu_n(\mathbb{R}_+^2) = \mu(\mathbb{R}_+^2)$. □

To prove the theorem, it remains to prove the absence of memory.

Lemma 5.3.5. *Let $t \in [0,1/\mathrm{m}[$, then $\big\{(T_i,G_i,R_i) \ : \ i\in\mathbb{N},\ T_i \leq t\big\}$ is independent of $\big\{(T_i,G_i,R_i) \ : \ i\in\mathbb{N},\ T_i > t\big\}$.*

Proof. First $\{(T_i, G_i, R_i) \ : \ T_i \leq t\}$ is given by $\{(t_i, l_i, x_i) : t_i \leq t, x_i \in [g(t), d(t)]\}$. Moreover $\{(T_i, G_i, R_i) \ : \ T_i > t\}$ depends on $(\mathcal{R}(t) - g(t)) \cap]-\infty, 0]$ and $\{(t_i, x_i - g(t), l_i) \ : \ t_i > t, \ x_i < g(t)\}$ which are independent. Moreover $(\mathcal{R}(t) - g(t)) \cap]-\infty, 0]$ is independent of $\{(t_i, l_i, x_i) : t_i \leq t, x_i \in [g(t), d(t)]\}$ by Lemma 5.1.2 and so is $\{(t_i, x_i - g(t), l_i) \ : \ t_i > t, \ x_i < g(t)\}$ by Poissonian property. This proves the result. $\square$

We can now prove the theorem and its corollary.

Proof of Theorem 5.3.1. We prove now that for every B finite union of disjoint rectangles of $[0, 1/\mathrm{m}[\times \mathbb{R}_+^2$:

$$\mathbb{P}(N_B = 0) = e^{-\gamma(B)}, \quad \text{where} \ \ \gamma(\mathrm{d}t\mathrm{d}y\mathrm{d}z) = \mathrm{d}t\rho^{(t)}(\mathrm{d}y\mathrm{d}z). \tag{5.6}$$

As γ is non atomic (use Lemma 5.3.3), this will ensure that $\{(T_i, G_i, R_i) : i \in \mathbb{N}\}$ is a PPP with intensity γ (use Rényi's Theorem [62]).

Let $t \in [0, 1/\mathrm{m}[$ and A a finite union of rectangles of $\mathbb{R}_+^2$. We consider $H(s) := \mathbb{P}(N_{]t,t+s] \times A} = 0)$ for $s \in [0, 1/\mathrm{m} - t[$. Lemma 5.3.5 entails that

$$H(s+h) = \mathbb{P}(N_{]t,t+s] \times A} = 0)\mathbb{P}(N_{]t+s,t+s+h] \times A} = 0) = H(s)\mathbb{P}(N_{]t+s,t+s+h] \times A} = 0).$$

We write $A = \sqcup_{i=1}^N A_i$ where A_i rectangle of $\mathbb{R}_+^2$. Theorem 5.2.1 and Lemma 5.3.4 ensure respectively that for all $1 \leq i, j \leq N$ such that $i \neq j$:

$$\lim_{h \to 0} \frac{\mathbb{P}(N_{]t,t+h] \times A_i} \geq 1, \ N_{]t,t+h] \times A_j} \geq 1)}{h} = 0 \quad ; \quad \lim_{h \to 0} \frac{\mathbb{P}(N_{]t,t+h] \times A_i} \geq 1)}{h} = \rho^{(t)}(A_i).$$

Then

$$\lim_{h \to 0} \frac{\mathbb{P}(N_{]t,t+h] \times A} \geq 1)}{h} = \sum_{i=1}^N \lim_{h \to 0} \frac{\mathbb{P}(N_{]t,t+h] \times A_i} \geq 1)}{h} = \rho^{(t)}(A),$$

and the derivative of H is given by

$$\lim_{h \to 0} \frac{H(s+h) - H(s)}{h} = H(s) \lim_{h \to 0} \frac{1 - \mathbb{P}(N_{]t+s,t+s+h] \times A} = 0)}{h} = H(s)\rho^{(t+s)}(A).$$

Thus $H(s)$ satisfies a differential equation of order 1 and we get (5.6) for $B =]t, t+s] \times A$.

$$H(s) = \exp\big(-\int_0^s \mathrm{d}u \rho^{(t+u)}(A)\big) = \exp\big(-\int_t^{t+s} \mathrm{d}u \rho^{(u)}(A)\big) = e^{-\gamma(]t,t+s] \times A)}$$

Using again Lemma 5.3.5 and additivity of measures proves (5.6) for every B finite union of rectangles of $[0, 1/\mathrm{m}[\times \mathbb{R}^+ \times \mathbb{R}^+$. $\square$

Proof of Corollary 5.3.2. As projection of the PPP $\{(T_i, G_i, R_i) : i \in \mathbb{N}\}$, $\{(T_i, R_i) : i \in N\}$ is a PPP with intensity $\mathrm{d}t \int_{y\in[0,\infty]} \rho^{(t)}(\mathrm{d}y\mathrm{d}z)$. By Lemma 5.3.3, we have :

$$\begin{aligned}
&\int_{y\in[0,\infty]} \rho^{(t)}(\mathrm{d}y\mathrm{d}z) \\
= \;& \mathrm{d}z\big(\bar{\nu}(z) + \int_z^\infty \nu(\mathrm{d}l) \int_0^\infty \mathbb{P}(\overleftarrow{\tau}^{(t)}_{l-z} \in \mathrm{d}x) \int_x^\infty \Pi^{(t)}(dy - x)(y - x)\big) \\
= \;& \mathrm{d}z\bar{\nu}(z)(1 + \int_0^\infty \Pi(\mathrm{d}y)y) \\
= \;& \mathrm{d}z \frac{\bar{\nu}(z)}{1 - \mathrm{m}t} \qquad \text{by (2.15)} \qquad (5.7)
\end{aligned}$$

which gives the intensity of $\{(T_i, R_i) : i \in \mathbb{N}\}$. In other words, $(R_i)_{i\in\mathbb{N}}$ is an iid sequence independent of $(T_i)_{i\in\mathbb{N}}$ such that $\mathbb{P}(R_i \in dz) = \mathrm{m}^{-1}\bar{\nu}(z)\mathrm{d}z,\ (z \geq 0)$. □

5.4 Evolution of $\mathbf{B}_0$

The processes $(g(t))_{t\in[0,1/\mathrm{m}]}$ and $(d(t))_{t\in[0,1/\mathrm{m}]}$ of the left and the right end-points of $\mathbf{B}_0$ have a quite different evolution, even though their one-dimensional distributions coincide. The process $(d(t))_{t\in[0,1/\mathrm{m}]}$ jumps each time $(g(t))_{t\in[0,1/\mathrm{m}]}$ jumps and each time a file arrives on $\mathbf{B}_0$. More precisely, there are two kinds of jumps of $(\mathbf{B}_0(t))_{t\in[0,1/\mathrm{m}]}$ corresponding respectively to :
- files which arrive at the left of $\mathbf{B}_0$ and cannot be entirely stored at its left (recall the previous section). These files induce the jumps $(-G_i, D_i)$ of the end-points of $\mathbf{B}_0$ at time T_i independently of the past (see Figure 4).
- files which arrive on $\mathbf{B}_0$. These files induce jumps of the right end-point $d(.)$ only, with total rate equal to $l(t)\bar{\nu}(0)$ (see Figure 5). This rate is infinite when $\bar{\nu}(0) = \infty$. Observe also that the jumps depend from the past of $\mathbf{B}_0$ through the value of the length $l(t)$.
Note that a file which arrives at the left of $\mathbf{B}_0(t-)$ at time t with remaining data of size R induces the same jump of the right end-point as a file of size R which arrives on $\mathbf{B}_0(t-)$ at time t. Obviously, the other files (files which are entirely stored at the left of $\mathbf{B}_0$ or which arrive at the right of $\mathbf{B}_0$) do not yield a jump of $\mathbf{B}_0$.

Thus, we define

$$D_i := d(T_i) - d(T_i^-)$$

and we decompose the process $(g(t), d(t))_{t\in[0,1/\mathrm{m}[}$ into two processes $(C^1(t))_{t\in[0,1/\mathrm{m}[}$ and $(C^2(t))_{t\in[0,1/\mathrm{m}[}$, which give the variation of the end-points of $\mathbf{B}_0$ respectively at times $(T_i)_{i\in\mathbb{N}}$ (due to the arrival of a file at the left of $g(t)$) and between successive times $(T_i)_{i\in\mathbb{N}}$ (due to the arrival of files on $\mathbf{B}_0(t)$). That is, for every $t \in [0, 1/\mathrm{m}[$,

$$C^1(t) := \sum_{T_i \le t} (-G_i, D_i), \qquad C^2(t) := \big(0, \sum_{\substack{0\le s\le t \\ s\notin\{T_i : i\in\mathbb{N}\}}} \Delta d(s)\big),$$

$$(g(t), d(t)) = C^1(t) + C^2(t).$$

First, we specify the distribution of $(C^1(t))_{t\in[0,1/\mathrm{m}]}$ (see below for the proofs).

Proposition 5.4.1. *The point process* $\big\{(T_i, G_i, D_i) : i \in \mathbb{N}\big\}$ *is a PPP on* $[0, 1/\mathrm{m}] \times \mathbb{R}_+^2$ *with intensity* $\mathrm{d}t\mu^{(t)}(\mathrm{d}y\mathrm{d}x)$*, where*

$$\mu^{(t)}(\mathrm{d}y\mathrm{d}x) = \int_0^\infty \rho^{(t)}(\mathrm{d}y\mathrm{d}z)\mathbb{P}(\overrightarrow{\tau}_z^{(t)} \in \mathrm{d}x).$$

We can now specify the distribution of the process $(g(t), d(t))_{t\in[0,1/\mathrm{m}[}$ as follows.

Theorem 5.4.2. $(g(t), d(t))_{t\in[0,1/\mathrm{m}[}$ *is a pure jump Markov process equal to* $(C^1(t) + C^2(t))_{t\in[0,1/\mathrm{m}]}$ *such that for all* $0 \le t \le t + s \le 1/\mathrm{m}$*,*

(i) $C^1(t + s) - C^1(t)$ *is independent of* $(g(u), d(u))_{u\in[0,t]}$.

(ii) Conditionally on $l(t) = l$*,* $C^2(t+s) - C^2(t)$ *is independent of* $(g(u), d(u))_{u\in[0,t]}$*. Conditionally also on* $T_i \le t \le t + s < T_{i+1}$ *for some* $i \in \mathbb{N}$ *:*

$$C^2(t + s) - C^2(t) \stackrel{d}{=} (0, \overrightarrow{\tau}_{S_{sl}}^{(t+s)}),$$

where $(S_x)_{x\ge 0}$ *is a subordinator with no drift and Lévy measure* ν*, which is independent of* $(\overrightarrow{\tau}_x^{(t+s)})_{x\ge 0}$*.*

Recalling that vague convergence of measures on A is the convergence of the integrals of measure against continuous functions with compact support in A, the jump rate of $(g(t), d(t))_{t\in[0,1/\mathrm{m}[}$ is then given by :

Corollary 5.4.3. *If* $t \in [0, 1/\mathrm{m}[$*, we have the following vague convergence of measures on* $[0, \infty[\times]0, \infty[$ *when* h *tends to* 0 *:*

$$h^{-1}\mathbb{P}(g(t) - g(t + h) \in \mathrm{d}y,\ d(t + h) - d(t) \in \mathrm{d}x \ \mid\ l(t) = l) \stackrel{v}{\Longrightarrow}$$
$$\mu^{(t)}(\mathrm{d}y\mathrm{d}x) + l\delta_0(\mathrm{d}y)\int_0^\infty \nu(\mathrm{d}z)\mathbb{P}(\overrightarrow{\tau}_z^{(t)} \in \mathrm{d}x).$$

We begin with two lemmas which state the independences needed for the proofs.

Lemma 5.4.4. $\{(T_i, G_i, D_i) : i \in \mathbb{N},\ T_i > t\}$ *is independent of* $(g(u), d(u))_{u\in[0,t]}$.

Proof. Using (5.8) below, we see that $\{(T_i, G_i, D_i) : i \in \mathbb{N},\ T_i > t\}$ is given by

$$\{(T_i, G_i, R_i) : i \in \mathbb{N}, T_i > t\} \quad \text{and} \quad (\overrightarrow{\mathcal{R}}(s))_{s>t}.$$

These quantities depend from the past through $(\overleftarrow{\mathcal{R}(t)}, \overrightarrow{\mathcal{R}(t)})$ which is independent of $(g(u), d(u))_{u\in[0,t]}$ by Proposition 5.1.1. □

Lemma 5.4.5. *Let* $i \in \mathbb{N}$ *and* $0 \le t' < t \le 1/\mathrm{m}$. *Conditionally on* $T_{i-1} = t'$ *and* $T_i = t$, $(\overrightarrow{\mathcal{R}(u)})_{u\in[t',t[}$ *is independent of the point process* $P_{g(t')}(t)$.

Proof. Conditioning by $T_{i-1} = t'$ and $T_i = t$ ensures that all the data arrived at the left of $g(t')$ during the time interval $[t', t[$ are stored at the left of $g(t')$. So $(\overrightarrow{\mathcal{R}(u)})_{u\in[t',t[}$ depdns only on the point process $P^{d(t')}_{g(t')}(t) \cup P^{d(t')}(t)$ which is independent of $P_{g(t')}(t)$ by Lemma 5.1.2. □

Proof of Proposition 5.4.1. At time T_i, the quantity of remaining data R_i is stored at the right of $\mathbf{B}_0(T_i-)$. It induces a jump $D_i = d(T_i) - d(T_i-)$ of the right end-point which is equal to R_i plus the sum of the lengths of blocks at the right of $\mathbf{B}_0(T_i-)$ which are reached during the storage of these data (see Figure 2). More precisely :

$$\begin{aligned} D_i &= \inf\{x \ge 0,\ \mid \mathcal{R}(T_i-) \cap [d(t), d(t)+x[| = R_i\} \\ &= \inf\{x \ge 0,\ \mid \overrightarrow{\mathcal{R}(T_i-)} \cap [0, x] \mid = R_i\} \\ &= \overrightarrow{\tau}^{(T_i-)}_{R_i}, \end{aligned} \tag{5.8}$$

by definition of $\overrightarrow{\tau}$ (see Section 3.3.1). Lemma 5.4.5 ensures that conditionally on $T_i = t$, $(\overrightarrow{\tau}^{(T_i-)}_x)_{x\ge0}$ is independent of (G_i, R_i) and distributed as $(\overrightarrow{\tau}^{(t)}_x)_{x\ge0}$. Then denoting by μ_t the law of (G_i, D_i) conditioned by $T_i = t$, we have

$$\mu_t(\mathrm{d}y\mathrm{d}x) = \mathbb{P}(G_t \in \mathrm{d}y,\ \overrightarrow{\tau}^{(t)}_{R_t} \in \mathrm{d}x), \tag{5.9}$$

where (G_t, R_t) is a random variable independent of $(\overrightarrow{\tau}^{(t)}_x)_{x\ge0}$ and distributed as (G_i, R_i) conditioned on $T_i = t$.

By Lemma 5.4.4, $\{(T_i, G_i, D_i) : i \in \mathbb{N},\ T_i > t\}$ is independent of $\{(T_i, G_i, D_i) : i \in \mathbb{N}, T_i \leq t\}$. Then conditionally on $(T_i)_{i\in\mathbb{N}}$, $(G_i, D_i)_{i\in\mathbb{N}}$ are independent. Adding that $\{T_i : i \in \mathbb{N}\}$ is a PPP on $[0, 1/\mathrm{m}]$ with intensity $\mathrm{d}t\mathrm{m}/(1 - \mathrm{m}t)$ ensures that $\{(T_i, G_i, D_i) : i \in \mathbb{N}\}$ is a (marked) PPP with intensity

$$\frac{\mathrm{m}}{1-\mathrm{m}t}\mathrm{d}t\mu_t(\mathrm{d}y\mathrm{d}x).$$

Further, by (5.9), this intensity is eqaul to

$$\mathrm{d}t\int_0^\infty \mathbb{P}(\overrightarrow{\tau}_z^{(t)} \in \mathrm{d}x)\frac{\mathrm{m}}{1-\mathrm{m}t}\mathbb{P}(G_t \in \mathrm{d}y,\ R_t \in \mathrm{d}z) = \mathrm{d}t\int_0^\infty \mathbb{P}(\overrightarrow{\tau}_z^{(t)} \in \mathrm{d}x)\rho^{(t)}(\mathrm{d}y\mathrm{d}z)$$

using Theorem 5.3.1. This completes the proof. □

Proof of Theorem 5.4.2.
(*i*) Thanks to Lemma 5.4.4, $C^1(t+s) - C^1(t)$ is independent of $(g(u), d(u))_{u\in[0,t]}$.

(*ii*) We condition by $T_i \leq t \leq t + s < T_{i+1}$ for some $i \in \mathbb{N}$ and $l(t) = l$. Then $g(t + s) - g(t) = 0$ and no data arrived at the left of $\mathbf{B}_0(t)$ during the time interval $]t, t+s]$ is stored at the right of this block. So the increment $d(t+s)-d(t)$ is caused by files arriving on $\mathbf{B}_0(t)$: they are stored at the right on $\mathbf{B}_0(t)$ and may join data already stored. Note that we can change the order of arrival of files between t and $t + s$ (use identity (3.4)). Thus, we first store the files which arrive at the right of $d(t)$ between times t and $t + s$, then the files which arrive on $\mathbf{B}_0(t)$ between times t and $t+s$ and we forget the files which arrive at the left of $g(t)$.

STEP 1 : At time t, we consider the half hardware at the right of $d(t)$ which we identify with $[0, \infty[$. Its free space is equal to $\overrightarrow{\mathcal{R}(t)}$. We store the files $i \in \{i \in \mathbb{N} : t_i \in]t, t + s], x_i > d(t)\}$ on this half hardware $[0, \infty[$ at location $x_i - d(t)$ following the process described in Introduction (the size of the file i is still l_i). Following Section 3.2, we get the counterpart of the characterization of the free space (3.4). That is, the new free space of the half hardware is equal to $\{x \geq 0 : \widetilde{Y}_x = \widetilde{I}_x\}$, where for every $x \geq 0$,

$$\widetilde{Y}_x = -x + \sum_{\substack{0\leq t_i\leq t+s \\ d(t)\leq x_i\leq d(t)+x}} l_i, \qquad \widetilde{I}_x := \inf\{\widetilde{Y}_y : 0 \leq y \leq x\}.$$

Using Lemma 5.1.2, we see that $\{(t_i, x_i - d(t), l_i) : x_i \geq d(t)\}$ is a PPP on $\mathbb{R}^{+3}$ with intensity $\mathrm{d}t \otimes \mathrm{d}x \otimes \nu(\mathrm{d}l)$. Then,

$$\left(\widetilde{Y}_x\right)_{x\geq 0} \stackrel{d}{=} \left(Y_x^{(t+s)}\right)_{x\geq 0}$$

is a Lévy process with Laplace exponent $\Psi^{(t+s)}$. As $[\Psi^{(t+s)}]'(0) < 0$, $\big(\widetilde{Y}_x\big)_{x\geq 0}$ is regular for $]-\infty, 0[$, in the sense that it takes negative values for some arbitrarily small x (Proposition 8 on page 84 in [19]). So for every stopping time T such that $\widetilde{Y}_T = \widetilde{I}_T$, there is the identity $T = \inf\{z \geq 0 : \widetilde{Y}_z < \widetilde{Y}_T\}$. This ensures that the free space $\{x \geq 0 : \widetilde{Y}_x = \widetilde{I}_x\}$ of the half hardware is the range of $(\widetilde{\tau}_x)_{x\geq 0}$ defined by

$$\widetilde{\tau}_x := \inf\{z \geq 0 : \widetilde{Y}_z < -x\}.$$

By Theorem 1 on page 189 in [19], $(\widetilde{\tau}_x)_{x\geq 0}$ is a subordinator with Laplace exponent $\kappa^{(t+s)}$, which is the inverse function of $-\Psi^{(t+s)}$. So $(\widetilde{\tau}_x)_{x\geq 0}$ is distributed as $(\overset{\rightarrow}{\tau}{}^{(t+s)}_x)_{x\geq 0}$. By Lemma 1 again, $\{(t_i, x_i - d(t), l_i) : x_i > d(t)\}$ is independent of $(g(u), d(u))_{u\in[0,t]}$. So $(\widetilde{\tau}_x)_{x\geq 0}$ is independent of $(g(u), d(u))_{u\in[0,t]}$.

STEP 2 : To obtain the covering $\mathcal{C}(t+s)$, we now store the files $\{i : t_i \in]t, t+s], x_i \in [g(t), d(t)[\}$. It amounts to store these files in the first free spaces (i.e. as much on the left as possible) of the half hardware considered above, whose free space is the range of $(\widetilde{\tau}_x)_{x\geq 0}$. The variation of the right end-point is equal to the sum of the sizes of these files, say S_t^{t+s}, plus the sizes of the lengths of the blocks of the half hardware joined during their storage. That is, as for (5.8),

$$C^2(t+s) - C^2(t) = (0, \widetilde{\tau}_{S_t^{t+s}}), \qquad \text{where} \quad S_t^{t+s} := \sum_{\substack{t < t_i \leq t+s \\ x_i \in [g(t), d(t)[}} l_i.$$

Conditionally on $l(t) = l$, by Poissonian property, $S_t^{t+s} \overset{d}{=} S_{sl}$, where $(S_x)_{x\geq 0}$ is a subordinator with no drift and Lévy measure ν. Adding that S_t^{t+s} is independent of $(\widetilde{\tau}_x)_{x\geq 0}$ gives the law of $C^2(t+s) - C^2(t)$. As $(\widetilde{\tau}_x)_{x\geq 0}$ and S_t^{t+s} are independent of $(g(u), d(u))_{u\in[0,t]}$, so is $C^2(t+s) - C^2(t)$.

These properties ensure that $(g(t), d(t))_{t\in[0,1/\mathrm{m}[}$ is a Markov process. $\square$

To prove Corollary 5.4.3, we need the following result which uses notation of Theorem 5.4.2.

Lemma 5.4.6. *We have the following vague convergence of measure on* $]0, \infty[$ *:*

$$h^{-1}\mathbb{P}\big(\overset{\rightarrow}{\tau}{}^{(t)}_{S_{hl}} \in \mathrm{d}x\big) \overset{v}{\Longrightarrow} l \int_0^\infty \nu(\mathrm{d}z)\mathbb{P}(\overset{\rightarrow}{\tau}{}^{(t)}_z \in \mathrm{d}x).$$

Proof. Denoting by ϕ the Laplace exponent of $(S_x)_{x\geq 0}$, $\big(\overset{\rightarrow}{\tau}{}^{(t)}_{S_{xl}}\big)_{x\geq 0}$ is a subordinator of Laplace exponent $l\phi \circ \kappa^{(t)}$ (see (**??**)). Moreover for every $\lambda \geq 0$, $\phi(\lambda) = \int_0^\infty (1 - e^{-\lambda y})\nu(\mathrm{d}y)$, which entails that

$$\phi \circ \kappa^{(t)}(\lambda) \quad = \quad \int_0^\infty \big(1 - e^{-z\kappa^{(t)}(\lambda)}\big)\nu(\mathrm{d}z)$$

$$= \int_0^\infty \mathbb{E}\big(1 - e^{-\lambda \vec{\tau}_z^{(t)}}\big)\nu(\mathrm{d}z)$$
$$= \int_0^\infty (1 - e^{-\lambda x}) \int_0^\infty \nu(\mathrm{d}z)\mathbb{P}(\vec{\tau}_z^{(t)} \in \mathrm{d}x).$$

Then $\big(\vec{\tau}^{(t)}_{S_{xl}}\big)_{x\geq 0}$ is a subordinator with no drift and Lévy measure

$$l \int_0^\infty \nu(\mathrm{d}z)\mathbb{P}(\vec{\tau}_z^{(t)} \in \mathrm{d}x).$$

Using Exercise 1 Chapter I in [19] or [20] on page 8 completes the proof. □

Proof of Corollary 5.4.3. We consider first the case when the increment of the left end-point is zero.

- Using Theorem 5.4.2 and recalling that $N_t^{t+h} = N_{]t+t+h]\times\mathbb{R}_+^2} = \mathrm{card}\{i \in \mathbb{N} : T_i \in]t, t+h]\}$, we have for all $c > 0$ such that $\int_0^\infty \nu(\mathrm{d}z)\mathbb{P}(\vec{\tau}_z^{(t)} = c) = 0$,

$$P(g(t+h) - g(t) = 0,\ d(t+h) - d(t) \geq c \ \mid\ l(t) = l) = \mathbb{P}(N_t^{t+h} = 0)\mathbb{P}(\vec{\tau}^{(t)}_{S_{hl}} \geq c). \tag{5.10}$$

Adding that $\mathbb{P}(N_t^{t+h} = 0) \xrightarrow{h\to 0} 1$ and using Lemma 5.4.6 give

$$h^{-1}P(g(t+h)-g(t) = 0,\ d(t+h)-d(t) \geq c \ \mid\ l(t) = l) \xrightarrow{h\to 0} l \int_0^\infty \nu(\mathrm{d}z)\mathbb{P}(\vec{\tau}_z^{(t)} \geq c). \tag{5.11}$$

- Let $a, b > 0$ and write

$$P(t, t+h) = \mathbb{P}(g(t) - g(t+h) \geq a,\ d(t+h) - d(t) \geq b \ \mid\ l(t) = l).$$

By Proposition 5.4.1, $\big\{(T_i, G_i, D_i) : i \in \mathbb{N}\big\}$ is a PPP on $[0, 1/\mathrm{m}]\times\mathbb{R}_+^2$ with intensity $\mathrm{d}t\mu^{(t)}(\mathrm{d}y\mathrm{d}x)$. The latter verifies $\mathbb{P}(N_t^{t+h} > 1) = o(h)$ $(h \to 0)$, so we have

$$h^{-1}\mathbb{P}(C^1(t+h) - C^1(t) \in]-\infty, -a] \times [b, \infty]) \xrightarrow{h\to 0} \mu^{(t)}([a, \infty[\times[b, \infty]). \tag{5.12}$$

We can prove now that

$$\lim_{h\to 0} h^{-1}P(t, t+h) = \mu^{(t)}([a, \infty[\times[b, \infty[). \tag{5.13}$$

- First we give the lower bound.

$$P(t, t+h) \geq \mathbb{P}(C^1(t+h) - C^1(t) \in]-\infty, -a] \times [b, \infty] \ \mid\ l(t) = l)$$

Using that $C^1(t+h) - C^1(t)$ is independent of $l(t)$ and (5.12), we get

$$\liminf_{h\to 0} h^{-1}P(t, t+h) \geq \mu^{(t)}([a, \infty[\times[b, \infty]). \tag{5.14}$$

- For the upper bound, observe that

$$\begin{aligned} P(t,t+h) &\leq \mathbb{P}(C^1(t+h)-C^1(t) \in]-\infty,-a] \times [b-\epsilon,\infty] \ \mid\ l(t)=l) \\ &+ \mathbb{P}(N_t^{t+h} \geq 1,\ C^2(t+h)-C^2(t) \in \{0\} \times [\epsilon,\infty[\ \mid\ l(t)=l). \end{aligned}$$

Using again $C^1(t+h)-C^1(t)$ is independent of $l(t)$ with (5.12) and Theorem 5.4.2 gives

$$\limsup_{h\to 0} h^{-1}P(t,t+h) \leq \mu^{(t)}([a,\infty[\times[b-\epsilon,\infty]).$$

Letting ϵ tend to 0 gives the upper bound :

$$\limsup_{h\to 0} h^{-1}P(t,t+h) \leq \mu^{(t)}([a,\infty[\times[b,\infty[).$$

The two limits (5.11) and (5.13) ensure the convergence of measures for sets of the form $\{0\} \times [c,d[$ (with $c>0$) and $[a,b[\times[c,d[$ (with $a>0$), which completes the proof. □

5.5 Evolution of the right end-point and of the length

Proposition 5.4.1, Theorem 5.4.2 and Corollary 5.4.3 give by projection :

Corollary 5.5.1. $(d(t))_{t\in[0,1/\mathrm{m}[}$ *is a jump process satisfying*

(i) $\big\{(T_i,D_i) : i\in\mathbb{N}\big\}$ *is a PPP on* $[0,1/\mathrm{m}[\times\mathbb{R}^+$ *with intensity*

$$\frac{\mathrm{d}t \int_{z\in[0,\infty]} \mathrm{d}z\bar{\nu}(z)\mathbb{P}(\overrightarrow{\tau}^{(t)}_z \in \mathrm{d}x)}{1-\mathrm{m}t},$$

and $\big\{(T_i,D_i) : i\in\mathbb{N},\ T_i>t\big\}$ *is independent of* $(d(u))_{u\in[0,t]}$.

(ii) For all $0\leq t\leq t+s<1/\mathrm{m}$ *:*
Conditionally on $l(t)=l$, $d(t+s)-d(t)$ *is independent of* $(d(u))_{u\in[0,t]}$.
Conditionally also on $T_i\leq t\leq t+s<T_{i+1}$ *for some* $i\in\mathbb{N}$ *:*

$$d(t+s)-d(t) \overset{d}{=} \overrightarrow{\tau}^{(t+s)}_{S_{sl}},$$

where $(S_x)_{x\geq 0}$ *is a subordinator with no drift and Lévy measure* ν, *that is independent of* $(\overrightarrow{\tau}^{(t+s)}_x)_{x\geq 0}$.

The jump rate of $(d(t))_{t\in[0,1/\mathrm{m}[}$ *is given by the following vague convergence of measures on* $]0,\infty[$ *for* h *tending to 0 :*

$$\frac{\mathbb{P}(d(t+h)-d(t)\in\mathrm{d}x \ \mid\ l(t)=l)}{h} \overset{v}{\Longrightarrow} \frac{\int_0^\infty \mathrm{d}z\bar{\nu}(z)\mathbb{P}(\overrightarrow{\tau}^{(t)}_z\in\mathrm{d}x)}{1-\mathrm{m}t} + l\int_0^\infty \nu(\mathrm{d}z)\mathbb{P}(\overrightarrow{\tau}^{(t)}_z\in\mathrm{d}x).$$

We stress that $(d(t))_{t\in[0,1/\mathrm{m}[}$ is not a Markov process since the jumps D_i before time t give informations about $l(t)$ and thus about the future of the process.
Note also that we can derive the law of $d(t)$ conditionally on $l(t)$ using Proposition 3.3.3. More precisely,

$$\forall d>0, \qquad \mathbb{P}(l(t)\in \mathrm{d}l \ \mid \ d(t)=d) = \mathbf{1}_{l\geq d}\frac{\Pi^{(t)}(\mathrm{d}l)}{\bar{\Pi}^{(t)}(d)}.$$

Finally we turn our interest to the process of the length $(l(t))_{t\in[0,1/\mathrm{m}]}$. Its increments which are due to files arrived at the left of $g(t)$ which are not stored entirely at the left $g(t)$, are denoted by L_i :

$$L_i := l(T_i) - l(T_i^-) = G_i + D_i.$$

The other increments of $(l(t))_{t\in[0,1/\mathrm{m}]}$ are due to files which arrive on $\mathbf{B}_0$. We can view $(l(t))_{t\in[0,1/\mathrm{m}]}$ as a branching process in continuous time with immigration L_i at time T_i (with no death, inhomogeneous branching and inhomogeneous immigration) :

Corollary 5.5.2. $(l(t))_{t\in[0,1/\mathrm{m}[}$ *is an inhomogeneous pure jump Markov process satisfying*

(i) $\{(T_i,L_i) : i\in\mathbb{N}\}$ *is a PPP on* $[0,1/\mathrm{m}[\times\mathbb{R}^+$ *with intensity*

$$\mathrm{d}t\int_{z\in[0,\infty]}\nu(\mathrm{d}z)\mathbb{P}(\overrightarrow{\tau}_z^{(t)}\in\mathrm{d}x)x,$$

and $\{(T_i,L_i) : i\in\mathbb{N}, T_i>t\}$ *is independent of* $(l(s))_{s\in[0,t]}$

(ii) Conditionally on $T_i\leq t\leq t+s<T_{i+1}$ *for some* $i\in\mathbb{N}$, $(l(t+u))_{u\in[0,t-s]}$ *satisfies the branching property : the law of* $(l(t+u))_{u\in[0,t-s]}$ *conditioned on* $l(t)=x+y$ *is equal to the law of the sum of two independent processes whose laws are respectively equal to* $(l(t+u))_{u\in[0,t-s]}$ *conditioned on* $l(t)=x$ *and* $(l(t+u))_{u\in[0,t-s]}$ *conditioned on* $l(t)=y$.

The jump rate of $(l(t))_{t\in[0,1/\mathrm{m}[}$ *is given by the following vague convergence of measures on* $]0,\infty[$ *for* h *tending to* 0 *:*

$$\frac{\mathbb{P}(l(t+h)-l(t)\in\mathrm{d}x \ \mid \ l(t)=l)}{h} \overset{v}{\Longrightarrow} (x+l)\int_0^{\infty}\nu(\mathrm{d}z)\mathbb{P}(\overrightarrow{\tau}_z^{(t)}\in\mathrm{d}x).$$

Example 5. For the basic example $\nu = \delta_1$, the jump rate of the length is equal to

$$\sum_{n=1}^{\infty} \frac{n+l}{n} e^{-tn} \frac{(tn)^{n-1}}{(n-1)!} \delta_n(\mathrm{d}x).$$

This is a consequence of the last displayed limit and (3.11).

Proof of Corollary 5.5.1. Using (5.7), we get :

$$\int_{z\in[0,\infty]} \mathbb{P}(\overrightarrow{\tau}_z^{(t)} \in \mathrm{d}x) \int_{y\in[0,\infty]} \rho^{(t)}(\mathrm{d}y\mathrm{d}z) = \frac{\int_{z\in[0,\infty]} \mathrm{d}z\bar{\nu}(z)\mathbb{P}(\overrightarrow{\tau}_z^{(t)} \in \mathrm{d}x)}{1-\mathrm{m}t},$$

which gives the intensity of $\{(T_i, D_i) : i \in \mathbb{N}\}$ by Proposition 5.4.1. □

Proof of Corollary 5.5.2. *(i)* Writing $L_i = G_i + D_i$, Proposition 5.4.1 entails that $\{(T_i, L_i) : i \in \mathbb{N}\}$ is a PPP on $[0, 1/\mathrm{m}] \times \mathbb{R}^+$ with intensity $\mathrm{d}t\widetilde{\mu}_t(\mathrm{d}x)$ where $\widetilde{\mu}_t$ is a measure on $\mathbb{R}^+$ defined for a Borel set A of $\mathbb{R}^+$ by

$$\widetilde{\mu}_t(A) = \int_{\mathbb{R}_+^2} \mathbb{1}_{\{y+y'\in A\}} \int_0^\infty \mathbb{P}(\overrightarrow{\tau}_z^{(t)} \in \mathrm{d}y')\rho^{(t)}(\mathrm{d}y\mathrm{d}z).$$

To determine $\widetilde{\mu}_t$, we compute its Laplace transform using Lemma 5.3.3 :

$$\begin{aligned}
\int_0^\infty e^{-\lambda x}\widetilde{\mu}_t(\mathrm{d}x) &= \int_{\mathbb{R}^{+3}} e^{-\lambda(y+y')}\rho^{(t)}(\mathrm{d}y\mathrm{d}z)\mathbb{P}(\overrightarrow{\tau}_z^{(t)} \in \mathrm{d}y') \\
&= \int_{\mathbb{R}^{+3}} e^{-\lambda y'}\mathbb{P}(\overrightarrow{\tau}_z^{(t)} \in \mathrm{d}y')\mathrm{d}z \int_z^\infty \nu(\mathrm{d}l)\big[e^{-\lambda y}\mathbb{P}(\overrightarrow{\tau}_{l-z}^{(t)} \in \mathrm{d}y) \\
&\qquad + \int_0^y e^{-\lambda x}\mathbb{P}(\overrightarrow{\tau}_{l-z}^{(t)} \in \mathrm{d}x)(y-x)e^{-\lambda(y-x)}\Pi^{(t)}(\mathrm{d}y-x)\big] \\
&= \int_0^\infty \mathrm{d}z e^{-z\kappa^{(t)}(\lambda)} \int_z^\infty \nu(\mathrm{d}l)e^{-(l-z)\kappa^{(t)}(\lambda)}\big[1 + \int_0^\infty e^{-\lambda u}u\Pi^{(t)}(\mathrm{d}u)\big] \\
&= \int_0^\infty \nu(\mathrm{d}l)le^{-l\kappa^{(t)}(\lambda)}[\kappa^{(t)}]'(\lambda) \\
&= -\frac{\partial}{\partial y}\bigg[\int_0^\infty \nu(\mathrm{d}l)e^{-l\kappa^{(t)}(y)}\bigg](\lambda) \\
&= -\frac{\partial}{\partial y}\bigg[\int_0^\infty e^{-yx}\int_0^\infty \nu(\mathrm{d}l)\mathbb{P}(\overrightarrow{\tau}_l^{(t)} \in \mathrm{d}x)\bigg](\lambda) \\
&= \int_0^\infty e^{-\lambda x}x\int_0^\infty \nu(\mathrm{d}l)\mathbb{P}(\overrightarrow{\tau}_l^{(t)} \in \mathrm{d}x).
\end{aligned}$$

Then $\widetilde{\mu}_t(\mathrm{d}x) = x\int_0^\infty \nu(\mathrm{d}z)\mathbb{P}(\overrightarrow{\tau}_z^{(t)} \in \mathrm{d}x)$, which gives the intensity of $\{(T_i, L_i) : i \in \mathbb{N}\}$.

(ii) The branching property can be seen as a consequence of the determination of the jump rate. We give here a more intuitive approach : We condition by $l(t) = x + y$ and by $T_i \leq t \leq t + s < T_{i+1}$ and we make the decomposition effective by splitting $\mathbf{B}_0(t)$ in two segments of length x and y. First we store the files $\{i : t_i \in]t, t+s], x_i > d(t)\}$. The free space of the half line at the right of $\mathbf{B}_0(t)$ is now the closed range a subordinator distributed like $(\overrightarrow{\tau}_x^{(t+s)})_{x\geq 0}$ (see STEP1 in the proof of Corollary 5.5.1). Then we store successively the files $\{i : t_i \in]t, t+s], x_i \in [g(t), g(t) + x]\}$ and $\{i : t_i \in]t, t+s], x_i \in]g(t) + x, d(t)]\}$ which induce two successive increments of the length. The free space at the right of 0 after the first storage keeps the same distribution and is independent of the first increment by strong regeneration. So the two increments are independent and distributed respectively like $l(t+s) - l(t)$ conditioned by $l(t) = x$ and by $l(t) = y$. This gives the result since $l(t)$ is Markovian. Formally $l(t+s) - l(t)$ is equal to $\overrightarrow{\tau}^{(t+s)}_{S_{s(x+y)}}$ (see proof of Proposition 5.4.1) and

$$\overrightarrow{\tau}^{(t+s)}_{S_{s(x+y)}} = \overrightarrow{\tau}^{(t+s)}_{S_{sx}} + \overrightarrow{\tau}^{(t+s)}_{S_{s(x+y)}} - \overrightarrow{\tau}^{(t+s)}_{S_{sx}}$$

gives the decomposition expected since $\overrightarrow{\tau}^{(t+s)}_{S_{s(x+y)}} - \overrightarrow{\tau}^{(t+s)}_{S_{sx}} \overset{d}{=} \overrightarrow{\tau}^{(t+s)}_{S_{sy}}$.

Using Corollary 5.4.3 and recalling the definition of $\widetilde{\mu}_t$ given at the beginning of the proof ensures that $h^{-1}\mathbb{P}(l(t+h) - l(t) \in \mathrm{d}x \mid l(t) = l)$ converges to

$$\widetilde{\mu}_t(\mathrm{d}x) + l\int_0^\infty \nu(\mathrm{d}z)\mathbb{P}(\overrightarrow{\tau}_z^{(t)} \in \mathrm{d}x).$$

The completes the proof, since $\widetilde{\mu}$ has been determined above. □

5.6 Complements

5.6.1 Distribution of $\{(T_i, G_i) : i \in \mathbb{N}\}$ derived from Theorem 5.3.1

In Section 5.4, we used the total intensity of the PPP $\{(T_i, G_i) : i \in \mathbb{N}\}$ to prove that the intensity of the PPP $\{(T_i, G_i, R_i) : i \in \mathbb{N}\}$ is equal to $\mathrm{d}t\rho^{(t)}(\mathrm{d}y\mathrm{d}z)$ (Theorem 5.3.1). Here we check that integrating this intensity with respect to the third coordinate enables us to recover the intensity of $\{(T_i, G_i) : i \in \mathbb{N}\}$ given in Theorem 5.2.1.
For that purpose, use Lemma 5.3.3 to rewrite $\rho^{(t)}$ as

$$\rho^{(t)}(\mathrm{d}y\mathrm{d}z) = \mathrm{d}z\int_0^\infty \nu(\mathrm{d}l + z)\big(\mathbb{P}(\overleftarrow{\tau}_l^{(t)} \in \mathrm{d}y) + \int_0^y \mathbb{P}(\overleftarrow{\tau}_l^{(t)} \in \mathrm{d}x)(y - x)\Pi^{(t)}(\mathrm{d}y - x)\big)$$

and calculate the Laplace transform of $\int_{z\in[0,\infty]}\rho^{(t)}(\mathrm{d}y\mathrm{d}z)$.

$$
\begin{aligned}
&\int_{y\in[0,\infty]} e^{-\lambda y}\int_{z\in[0,\infty]}\rho^{(t)}(\mathrm{d}y\mathrm{d}z)\\
=\;&\int_0^\infty\int_0^\infty \mathrm{d}z\nu(\mathrm{d}l+z)\int_0^\infty e^{-\lambda y}\big[\mathbb{P}(\overleftarrow{\tau}_l^{(t)}\in dy)+\int_0^y \mathbb{P}(\overleftarrow{\tau}_l^{(t)}\in \mathrm{d}x)(y-x)\Pi^{(t)}(\mathrm{d}y-x)\big]\\
=\;&\int_0^\infty \mathrm{d}l\bar{\nu}(l)\big[e^{-l\kappa(\lambda)}+\int_0^\infty \mathbb{P}(\overleftarrow{\tau}_l^{(t)}\in \mathrm{d}x)e^{-\lambda x}\int_x^\infty e^{-\lambda(y-x)}(y-x)\Pi^{(t)}(\mathrm{d}y-x)\big]\\
=\;&\int_0^\infty \mathrm{d}l\bar{\nu}(l)e^{-l\kappa(\lambda)}[\kappa^{(t)}]'(\lambda)\\
=\;&\int_0^\infty \mathrm{d}l\frac{\bar{\nu}(l)}{l}\frac{\partial}{\partial\lambda}\mathbb{E}(-e^{-l\kappa^{(t)}(\lambda)})\\
=\;&\int_0^\infty \mathrm{d}l\frac{\bar{\nu}(l)}{l}\frac{\partial}{\partial\lambda}\mathbb{E}(-e^{-\lambda\overleftarrow{\tau}_l^{(t)}})\\
=\;&\int_0^\infty \mathrm{d}l\frac{\bar{\nu}(l)}{l}\int_0^\infty e^{-\lambda y}y\mathbb{P}(\overleftarrow{\tau}_l^{(t)}\in \mathrm{d}y)\\
=\;&\int_0^\infty \mathrm{d}ye^{-\lambda y}\int_0^\infty \mathbb{P}(Y_y^{(t)}\in -\mathrm{d}l)\bar{\nu}(l)\quad\text{using (3.9).}
\end{aligned}
$$

Thus, we conclude with

$$\mathrm{d}t\int_{z\in[0,\infty]}\rho^{(t)}(\mathrm{d}y\mathrm{d}z)=\mathrm{d}t\mathrm{d}x\int_0^\infty \mathbb{P}(Y_x^{(t)}\in -\mathrm{d}l)\bar{\nu}(l).$$

5.6.2 Direct proof of Corollary 5.3.2 using fluctuation theory

Here we determine the distribution of the remaining data using fluctuation theory : we get laws at fixed times and do not need Theorem 5.2.1, as for the proof of Section 5.4.

We fix t,h and $x\geq 0$. We add the lengths of files fallen in $[g(t)-x,g(t)]$ during the time interval $]t,t+h]$. Then we remove the free space in $[g(t)-x,g(t)]$ at time t which is equal to $L_x^{(t)}$. The sum of data arrived at the left of $\mathbf{B}_0(t)$ not stored at the left of $\mathbf{B}_0(t)$ between time t and $t+h$ is equal to the maximum in $x\geq 0$ of this difference. It is also the quantity of data which has tried to occupy the location $g(t)$ (successfully or not) between time t and $t+h$: $Y_{g(t)}^{(t+h)}-I_{g(t)}^{(t+h)}$. So, we have

Lemma 5.6.1. *Let $0\leq t<1/\mathrm{m}$ and $h\geq 0$, then*

$$Y_{g(t)}^{(t+h)}-I_{g(t)}^{(t+h)}=\sup\{S_{hx}-L_x^{(t)},x\geq 0\}=\sup\{S_{h\overleftarrow{\tau}_x^{(t)}}-x,x\geq 0\}\qquad a.s,$$

where $(S_x)_{x\geq 0}$ is a subordinator with drift $\mathrm{d}=0$ and Lévy measure $\nu(dx)$, which is independent of $(L_x^{(t)})_{x\geq 0}$ and $(\overleftarrow{\tau}_x^{(t)})_{x\geq 0}$.

Denoting $S^{(t,h)} := \sup\{S_{h\overleftarrow{\tau}^{(t)}_x} - x, x \geq 0\}$, we have for all $0 < a \leq b$,

$$\lim_{h\to 0} h^{-1}\mathbb{P}(S^{(t,h)} \in [a,b]) = \lim_{h\to 0} h^{-1}\mathbb{P}(\exists i \in \mathbb{N} : (T_i, R_i) \in]t, t+h] \times [a,b])$$

and we find the law given in Corollary 5.3.2 :

Proposition 5.6.2. *We have the following weak convergence of bounded measures on* $]0,\infty[$ *when* h *tends to* 0 *:*

$$\frac{\mathbb{P}(S^{(t,h)} \in \mathrm{d}x)}{h} \overset{w}{\Longrightarrow} \frac{\bar{\nu}(x)\mathrm{d}x}{1-\mathrm{m}t}.$$

Proof. $(S_{h\overleftarrow{\tau}^{(t)}_x} - x)_{x\geq 0}$ is a lévy process with negative drift -1, no negative jumps and bounded variation. Its Laplace exponent is $\kappa^{(t)} \circ (h\phi) - id$, where ϕ is the Laplace exponent of S and is defined by

$$\forall \lambda \geq 0, \quad \phi(\lambda) = \int_0^\infty (1 - e^{-\lambda x})\nu(\mathrm{d}x).$$

Note also that using (2.15), we have

$$\big[\kappa^{(t)} \circ (h\phi) - id\big]'(0) = [\kappa^{(t)}]'(0).h.\phi'(0) - 1 = \frac{1}{1-\mathrm{m}t}\mathrm{m}h - 1, \tag{5.15}$$

which is negative since $0 \leq t + h < 1/\mathrm{m}$. Then identity (14) in [13] or Theorem 5 in [19] ensure that $\forall \lambda > 0$, $\forall h \in [0, 1/\mathrm{m} - t[$,

$$\mathbb{E}\big(\exp(-\lambda S^{(t,h)})\big) = \left(\frac{1}{1-\mathrm{m}t}\mathrm{m}h - 1\right)\frac{\lambda}{(\kappa^{(t)} \circ (h\phi) - id)(\lambda)}$$

Moreover,

$$\frac{(\kappa^{(t)} \circ (h\phi) - id)(\lambda)}{\lambda} = \frac{\kappa^{(t)}(h\phi(\lambda))}{h\phi(\lambda)}\frac{h\phi(\lambda)}{\lambda} - 1 = -1 + \frac{1}{1-\mathrm{m}t}\frac{h\phi(\lambda)}{\lambda} + \circ_{h\to 0}(h).$$

So

$$\mathbb{E}\big(\exp(-\lambda S^{(t,h)})\big) = 1 + \frac{1}{1-\mathrm{m}t}\big(\frac{\phi(\lambda)}{\lambda} - \mathrm{m}\big)h + \circ_{h\to 0}(h).$$

We can now prove the convergence of $h^{-1}\mathbb{P}(S^{(t,h)} > x)$ when h tends to 0.

$$\lim_{h\to 0}\int_0^\infty e^{-\lambda x}\frac{\mathbb{P}(S^{(t,h)} > x)}{h}\mathrm{d}x = \lim_{h\to 0}\frac{1 - \mathbb{E}\big(\exp(-\lambda S^{(t,h)})\big)}{h\lambda} = \frac{1}{1-\mathrm{m}t}\big(\frac{\mathrm{m}}{\lambda} - \frac{\phi(\lambda)}{\lambda^2}\big).$$

Moreover Fubini gives

$$\int_0^\infty \mathrm{d}x e^{-\lambda x} \int_x^\infty \bar{\nu}(a)\mathrm{d}a = \int_0^\infty \nu(\mathrm{d}y) \int_0^y \mathrm{d}a \frac{1-e^{-\lambda a}}{\lambda} = \frac{\mathrm{m}}{\lambda} - \frac{\phi(\lambda)}{\lambda^2}.$$

Then for every $\lambda > 0$,

$$\lim_{h\to 0} \int_0^\infty e^{-\lambda x} \frac{\mathbb{P}(S^{(t,h)} > x)}{h} \mathrm{d}x = \int_0^\infty e^{-\lambda x} \frac{\int_x^\infty \bar{\nu}(a)\mathrm{d}a}{1-\mathrm{m}t} \mathrm{d}x,$$

which proves the convergence of $\mathbb{P}(S^{(t,h)} \in \mathrm{d}x)/h$ to $\bar{\nu}(x)\mathrm{d}x/(1-\mathrm{m}t)$. Indeed, introduce the measures $\mu_h(\mathrm{d}x)$ and $\mu(\mathrm{d}x)$ on $\mathbb{R}^+$ whose tails are given by

$$\mu_h(]x,\infty]) = e^{-x}\mathbb{P}(S^{(t,h)} > x)/h, \qquad \mu(]x,\infty]) = e^{-x} \int_x^\infty \bar{\nu}(a)\mathrm{d}a/(1-\mathrm{m}t).$$

The last displayed limit entails the weak convergence of $\mu_h(\mathrm{d}x)$ to $\mu(\mathrm{d}x)$ when h tends to 0, by convergence of Laplace transforms. As μ is non atomic, for every $x \geq 0$, $\mu_h(]x,\infty])$ tends to $\mu(]x,\infty])$, which proves that $\mathbb{P}(S^{(t,h)} > x)/h$ tends to $\int_x^\infty \bar{\nu}(a)\mathrm{d}a/(1-\mathrm{m}t)$. $\square$

Remark 10. Denote $\gamma^{(t,h)}$ the a.s instant at which the supremum $S^{(t,h)}$ is reached. To obtain the distribution of $\{(T_i, G_i, R_i) \; : \; i \in \mathbb{N}\}$ by this way, we need to know the joint law of $(S^{(t,h)}, \overleftarrow{\tau}^{(t)}_{\gamma^{(t,h)}})$ which we cannot derive directly from fluctuation theory.

Part II

Processus de branchement pour l'infection de parasites

Chapter 6

Background on Branching Processes in Random Environment

We gather here some important results about Branching Process in Random Environment (BPRE). We consider a BPRE $(Z_n)_{n\in\mathbb{N}}$ specified by a sequence of iid generating functions $(f_n)_{n\in\mathbb{N}}$ distributed as f [6, 7, 43]. More precisely, conditionally on the environment $(f_n)_{n\in\mathbb{N}}$, particles at generation n reproduce independently of each other and their offspring sizes have generating function f_n. Then Z_n is the number of particles at generation n and Z_{n+1} is the sum of Z_n independent random variables with generating function f_n. That is, for every $n \in \mathbb{N}$,

$$\mathbb{E}\big(s^{Z_{n+1}}|Z_0, \dots, Z_n;\ f_0, \dots, f_n\big) = f_n(s)^{Z_n} \qquad (0 \leq s \leq 1).$$

Thus, denoting by $F_n := f_0 \circ \cdots \circ f_{n-1}$, we have for every $k \in \mathbb{N}$,

$$\mathbb{E}_k(s^{Z_{n+1}} \mid\ f_0, ...,\ f_n) = \mathbb{E}(s^{Z_{n+1}} \mid Z_0 = k,\ f_0, ...,\ f_n) = F_n(s)^k \qquad (0 \leq s \leq 1).$$

When the environments are deterministic (i.e. f is a deterministic generating function), this process is the Galton Watson process (GW) with reproduction law Z, where f is the generating function of Z.

The process is called subcritical, critical or supercritical respectively if

$$\mathbb{E}\big(\log(f'(1))\big)$$

is negative, zero or positive. This process becomes extinct a.s. :

$$\mathbb{P}(\exists n \in \mathbb{N} : Z_n = 0) = 1$$

iff it is subcritical or critical [6, 8].

6.1 Subcritical BPRE

For a subcritical GW process, if $\mathbb{E}(Z_1 \log^+(Z_1)) < \infty$, recall that there exists $c > 0$ such that $\mathbb{P}(Z_n > 0) \sim cf'(1)^n$ when n tends to infinity (see [8]). In random environments, this asymptotic behavior depends on whether the BPRE is strongly subcritical (SS), intermediate subcritical (IS) or weakly subcritical (WS), as stated below. Moreover a subcritical GW process is strongly subcritical (SS).

Note that $s \in \mathbb{R}^+ \mapsto \mathbb{E}(f'(1)^s)$ is a convex function and define γ and α in $[0, 1]$ such that

$$\gamma := \inf_{\theta \in [0,1]} \big\{\mathbb{E}\big(f'(1)^\theta\big)\big\} = \mathbb{E}\big(f'(1)^\alpha\big). \tag{6.1}$$

From now on, we assume $\mathbb{E}(f'(1)|\log(f'(1))|) < \infty$. Note also that $0 < \gamma < 1$, $\gamma \leq \mathbb{E}(f'(1))$ and

$$\gamma = \mathbb{E}(f'(1)) \;\Leftrightarrow\; \mathbb{E}\big(f'(1)\log(f'(1))\big) \leq 0.$$

There are three subcases (see [43]).

- ⋆ The strongly subcritical case (SS), where $\mathbb{E}(f'(1)\log(f'(1))) < 0$. In this case, assuming further
$$\mathbb{E}(Z_1 \log^+(Z_1)) < \infty,$$
then there exist $c, \alpha_k > 0$ such that, as $n \to \infty$:
$$\mathbb{P}_k(Z_n > 0) \sim c\alpha_k \mathbb{E}(f'(1))^n, \qquad \alpha_1 = 1. \tag{6.2}$$

- ⋆ The intermediate subcritical case (IS), where $\mathbb{E}(f'(1)\log(f'(1))) = 0$. In this case, assuming further
$$\mathbb{E}\big(f'(1)\log^2(f'(1))\big) < \infty, \qquad \mathbb{E}\big([1 + \log^-(f'(1))]f''(1)\big) < \infty,$$
then there exist $c, \alpha_k > 0$ such that as $n \to \infty$:
$$\mathbb{P}_k(Z_n > 0) \sim c\alpha_k n^{-1/2}\mathbb{E}(f'(1))^n, \qquad \alpha_1 = 1. \tag{6.3}$$

- ⋆ The weakly subcritical case (WS), where $0 < \mathbb{E}(f'(1)\log(f'(1))) < \infty$. In this case, assuming further
$$\mathbb{E}(f''(1)/f'(1)^{1-\alpha}) < \infty, \qquad \mathbb{E}(f''(1)/f'(1)^{2-\alpha}) < \infty,$$
then there exist $c, \alpha_k > 0$ such that as $n \to \infty$:
$$\mathbb{P}_k(Z_n > 0) \sim c\alpha_k n^{-3/2}\gamma^n, \qquad \alpha_1 = 1. \tag{6.4}$$

In the next chapter, for each case, we take the integrability assumptions above for granted. See [91] for asymptotics with weaker hypothesis in the (IS) case.

It is also known that the process Z_n starting from k particles and conditioned to be non zero converges to a finite positive random variable Υ_k, called the Yaglom quasistationary distribution (see [43]) :

$$\mathbb{E}_k\big(s^{Z_n}|\ Z_n > 0\big) \stackrel{n\to\infty}{\longrightarrow} \mathbb{E}\big(s^{\Upsilon_k}\big).$$

See Section 3.3 for discussions about $(\Upsilon_k)_{k\in\mathbb{N}}$.

Actually, in [43], the result and the proof of these convergences are given for $k=1$. They can be generalized to $k\geq 1$ with the following modifications. Set

$$S_n := \log(k)+\log(f_0'(1))+...+\log(f_{n-1}'(1)), \quad g_0(s) := \frac{1}{1-f_0(s)^k}-\frac{1}{kf_0'(1)(1-s)},$$

and recalling Notations of [43]

$$f_{k,l} := \begin{cases} f_k\circ f_{k+1}\circ\cdots\circ f_{l-1}, & k<l \\ f_{k-1}\circ f_{k-2}\circ\cdots\circ f_l, & k>l \\ \mathrm{id}, & k=l. \end{cases}$$

Then $1-\mathbb{E}_k(s^{Z_n}|Z_n>0) = \mathbb{E}(1-f_{0,n}^k(s))/\mathbb{P}_k(Z_n>0)$. Lemma 2.1 of [43] still holds replacing $f_{0,n}$ by $f_{0,n}^k$ and $\mathbb{P}(Z_n>0)$ by $\mathbb{P}_k(Z_n>0)$. Lemma 2.2 also still holds and results of Lemma 2.3. can now be stated as follows. By convexity of $x\in[0,1]\to x^k$ and $(f_n)_{n\in\mathbb{N}}$, for every $n\geq 0$, we have $a.s$ $\exp(-S_i)(1-f_{i,0}(s))\leq 1$ $(0\leq s\leq 1)$. Moreover $\exp(-S_n))(1-f_{n,0}(s)^k)$ converges $a.s.$ as $n\to\infty$, which is a direct consequence of the convergence for $k=1$ given in Lemma 2.3. in [43] (noting also that this implies $f_{n,0}(s)\to 0$ $a.s.$ as $n\to\infty$).

Finally, let us label by $i\in\mathbb{N}$ the initial particles and denote by $Z_n^{(i)}$ the number of descendants of particle i at generation n.

We consider the case where reproduction laws are a.s. linear fractional, since in that case survival probabilities can be computed explicitly. Thus, there are two random variables $A\in[0,\infty[$ and $B\in[0,1[$ with $A+B\leq 1$ such that

$$f(s) = 1-\frac{A}{1-B}+\frac{As}{1-Bs} \quad a.s. \qquad (0\leq s\leq 1). \tag{6.5}$$

In this case, setting for every $i\in\mathbb{N}$,

$$P_i := f_{n-i}'(1)...f_{n-1}'(1), \qquad (P_0=1),$$

we have (see [4], [46] or [63])

$$\mathbb{P}(Z_n > 0 | Z_0 = 1,\ f_0, ...,\ f_{n-1}) = 1 - F_n(0) = \left(1 + \sum_{i=0}^{n-1} \frac{f''_{n-i-1}(1)}{2f'_{n-i-1}(1)} P_i\right)^{-1} P_n. \quad (6.6)$$

Considering $((Z_n^{(i)})_{n\in\mathbb{N}},\ i \geq 1)$ such that conditionally on $(f_0, ..., f_{n-1})$, $(Z_n^{(i)},\ i \geq 1)$ is an iid sequence with common p.g.f F_n, we get

$$\mathbb{P}(Z_n^{(1)} > 0, ..., Z_n^{(k)} > 0 |\ f_0, ...,\ f_{n-1}) = \left(1 + \sum_{i=0}^{n-1} \frac{f''_{n-i-1}(1)}{2f'_{n-i-1}(1)} P_i\right)^{-k} P_n^k. \quad (6.7)$$

For a general BPRE, we use now that for every probability generating function f_i, we can find $\widetilde{f}_i$ linear fractional probability generating function such that for every $s \in [0, 1]$, $\widetilde{f}_i(s) \geq f_i(s)$, $\widetilde{f}'_i(1) = f'_i(1)$, $\widetilde{f}''_i(1) = 2f''_i(1)$ (see [46] or [63]). Then, $\widetilde{F}_n(0) \geq F_n(0)$ $a.s.$ ensures that

$$\mathbb{P}(Z_n > 0 \mid f_0, ..., f_{n-1}) \geq \mathbb{P}(\widetilde{Z}_n > 0 \mid f_0, ..., f_{n-1}) \qquad a.s. \quad (6.8)$$

More generally, for every $k \geq 1$,

$$\begin{aligned} & \mathbb{P}(Z_n^{(1)} > 0,\ Z_n^{(2)} > 0,\ ...,\ Z_n^{(k)} > 0 \mid f_0, ..., f_{n-1}) \\ = \ & (1 - F_n(0))^k \\ \geq \ & (1 - \widetilde{F}_n(0))^k \\ = \ & \mathbb{P}(\widetilde{Z}_n^{(1)} > 0,\ \widetilde{Z}_n^{(2)} > 0,\ ...,\ \widetilde{Z}_n^{(k)} > 0 \mid f_0, ..., f_{n-1}) \qquad a.s. \quad (6.9) \end{aligned}$$

6.2 Critical BPRE

In the critical case, for a GW process, there exists $d > 0$ such that (see [8])

$$\mathbb{P}(Z_n > 0) \sim d/\sqrt{n} \qquad (n \to \infty).$$

For a critical BPRE, under the following moment assumption

$$0 < \mathbb{E}(\log(f'_0(1))^2) < \infty, \quad \mathbb{E}\big([1 + \log(f'_0(1))]f''_0(1)/2f'_0(1)\big) < \infty,$$

there exist $0 < c_1 < c_2 < \infty$ such that for every $n \in \mathbb{N}$ (see [63])

$$c_1/\sqrt{n} \leq \mathbb{P}(Z_n > 0) \leq c_2/\sqrt{n}.$$

Kozlov [63] has also proved that, in the linear fractional case, there exists $d > 0$ such that

$$\mathbb{P}(Z_n > 0) \sim d/\sqrt{n} \qquad (n \to \infty).$$

See also [3] for the existence of slowing varying function l such that

$$\mathbb{P}_1(Z_n > 0) \sim l(n)/\sqrt{n}, \qquad (n \to \infty),$$

in a more general context.

6.3 Supercritical BPRE

First we have the following expected result in the supercritical case [7, 8].

Proposition 6.3.1. *If* $\mathbb{E}(\log(f'(1))) > 0$, $\mathbb{P}(Z_n \overset{n\to\infty}{\longrightarrow} \infty \mid \forall n \in \mathbb{N} : Z_n > 0) = 1$.

Moreover, we have the following analogue of Kesten-Stigum theorem. Assuming that

$$\mathbb{E}\big(f'(1)\sum_{k=0}^{\infty} k\log(k) f^{(k)}(0)/k!\big) < \infty,$$

we have

$$W_n = Z_n/(\Pi_{i=0}^{n-1} f_i'(1)) \overset{n\to\infty}{\longrightarrow} W, \quad \mathbb{P}(W > 0) = p$$

where $p > 0$ is the survival probability of the BPRE.

Tail of W at 0. In the case $\mathbb{P}(f(0) = 0) = 0$ and f is not a.s. the identity, we have the following results (see [51]). Define

$$\alpha = -\frac{\mathbb{E}[\log(f'(0))]}{\mathbb{E}[\log(f'(1))]}, \qquad \gamma = \frac{\mathbb{E}[\log(\inf\{k \in \mathbb{N}^* : f^{(k)}(0) > 0\})]}{\mathbb{E}[\log(f'(1))]}.$$

In the Böttcher case ($\mathbb{P}(Z_1 \leq 1) = 0$), for all $\epsilon > 0$, there exist $a, b, c, d > 0$ such that

$$a\exp\big(-b\delta^{(\gamma+\epsilon)/(\gamma-1-\epsilon)}\big) \leq \mathbb{P}(W < \delta) \leq c\exp\big(-d\delta^{(\gamma+\epsilon)/(\gamma-1-\epsilon)}\big), \quad (\delta > 0).$$

In the Schröder case($\mathbb{P}(Z_1 = 1) > 0$ and $\mathbb{P}(Z_1 = 0) = 0$), for every $\epsilon > 0$, there exist $a, b > 0$ such that

$$a\delta^{\alpha+\epsilon} \leq \mathbb{P}(W < \delta) \leq b\delta^{\alpha+\epsilon}, \quad (\delta > 0).$$

Moments By Theorem 3 in [46], we have

Proposition 6.3.2. *For every* $s > 1$*,* W_n *converges to* W *in* L^s *iff*

$$\mathbb{E}(f'(1)^{1-s}) < 1, \qquad \mathbb{E}\big((Z_1/f'(1))^s\big) < \infty.$$

As a consequence, W_n converges to W for all $s > 1$ and all moments of W are finite iff

$$f'(1) \geq 1 \text{ a.s.}; \quad \mathbb{P}(f'(1) > 1) > 0, \quad \forall s > 1,\ \mathbb{E}\big((Z_1/f'(1))^s\big) < \infty.$$

That is, the environments are a.s. critical or supercritical and they are supercritical with a positive probability.

Tail of W in ∞. If

$$\mathbb{P}(f'(1) < 1) > 0, \qquad \mathbb{P}(f'(1) > 1) > 0, \qquad \forall k \in \mathbb{N}, \mathbb{E}(f'(1)^{-k}) < \infty,$$

then there exists $\xi > 1$ such that

$$\mathbb{E}(f'(1)^{1-\xi}) = 1,$$

and, by Theorem 2.2 in [66],

$$0 < \liminf_{x\to\infty} x^{\xi}\mathbb{P}(W > x) \leq \liminf_{x\to\infty} x^{\xi}\mathbb{P}(W > x) < \infty.$$

Chapter 7

Limit theorems for subcritical Branching Processes in Random environment

7.1 Introduction

We consider a Branching Process in Random Environment (BPRE) $(Z_n)_{n\in\mathbb{N}}$ specified by a sequence of iid generating functions $(f_n)_{n\in\mathbb{N}}$ distributed as f [2, 6, 7, 43]. That is, for every $n \in \mathbb{N}$,

$$\mathbb{E}\big(s^{Z_{n+1}}|Z_0,\dots,Z_n;\ f_0,\dots,f_n\big) = f_n(s)^{Z_n} \qquad (0 \le s \le 1).$$

Recall also that, denoting by $F_n := f_0 \circ \cdots \circ f_{n-1}$, we have for every $k \in \mathbb{N}$,

$$\mathbb{E}_k(s^{Z_{n+1}} \mid\ f_0,\dots,\ f_n) = \mathbb{E}(s^{Z_{n+1}} \mid Z_0 = k,\ f_0,\dots,\ f_n) = F_n(s)^k \qquad (0 \le s \le 1).$$

In this paper, we consider the subcritical case :

$$\mathbb{E}\big(\log(f'(1))\big) < 0.$$

This is the case where extinction occurs a.s., that is

$$\mathbb{P}(\exists n \in \mathbb{N} : Z_n = 0) = 1.$$

Recall that a subcritical BPRE can be strongly subcritical (SS), intermediate subcritical (IS) or weakly subcritical (WS) (see [43] or the previous Section for details) (see the previous chapter).

We study the role of the initial number of particles in the limit theorems. For a GW process, particles are independent. As a consequence, limit theorems starting with several particles can be directly derived from the case with one single initial particle.

In random environment, particles do not reproduce independently; more precisely independence holds only conditionally on the environments. This explains why asymptotics for (WS) BPRE starting with several particles are different from the analogous results for a GW process.
When the BPRE is (SS) or (IS), conditioning on the survival of the population at generation n, only one initial particle survives in generation n when $n \to \infty$, just as for a GW process. But this does not hold in the (WS) case (see forthcoming Proposition 7.2.3). Thus, (WS) BPRE conditioned to survive have a supercritical behavior, as previously observed in [2].
We give an interpretation of these results in terms of environments (see Section 7.2.3 for details). Particles die out a.s. and we want to explain what happened when the population survives a long time. Roughly speaking, we prove that in the (SS+IS) case, survival of the population is due to exceptional multiplication of particles in 'bad environments', whereas in the (WS) case, survival of particles is due to exceptionally 'nice environments' (and normal multiplication in these environments). More precisely, conditioning on non-extinction induces a selection of environments with high reproduction law. In the (SS+IS) case, we prove that the survival probability of the branching process in the environments selected is still zero. This is obvious if environments are a.s. subcritical, i.e. $f'(1) < 1$ a.s. But in the (WS) case, conditioning by the survival of the population select only supercritical environments. That is, the sequence of environments selected has a.s. a positive survival probability (Theorem 7.2.5). Finally we make the initial number of particles tend to infinity and the sequence of environments becomes subcritical again.

We determine how the asymptotic survival probability depends on the initial number of particles. In that view, we define

$$\alpha_k := \lim_{n\to\infty} \mathbb{P}_k(Z_n > 0)/\mathbb{P}_1(Z_n > 0).$$

For a GW process, $\alpha_k = k$. That is, the asymptotic survival probability is proportional to the initial number of particles. This equality still holds in the (SS+IS) case for BPRE, but not in the (WS) case where a different asymptotic as $k \to$ tends to infinity is established (see forthcoming Theorem 7.2.2). For the proof, we need an asymptotic result on random walks with negative drift (Section 7.4), which gives the product of the means of the successive environments.
In the supercritical case, see [45] for asymptotics of the extinction probability when the number of initial particles tends to infinity.

In Section 7.2.4, we are interested in the characterization of the Yaglom quasistationary distribution, that is the limit as $n \to \infty$ of the number of particles at generation n, conditioned to be nonzero, starting with k particles.

Finally, in Section 7.2.5, we focus on the Q-process associated to the subcritical

BPRE, which is defined for all $l_1, l_2, ..., l_n \in \mathbb{N}$, by

$$\mathbb{P}_k(Y_1 = l_1, ..., Y_n = l_n) = \lim_{p\to\infty} \mathbb{P}_k(Z_1 = l_1, ..., Z_n = l_n | Z_{n+p} > 0).$$

See [8] for details on the Q-process associated to GW.

Recalling results given in the previous Chapter, we can define γ and α in $[0,1]$ such that

$$\gamma := \inf_{\theta\in[0,1]} \{\mathbb{E}(f'(1)^\theta)\} = \mathbb{E}(f'(1)^\alpha). \tag{7.1}$$

Finally, we recall asymptotics in the different cases, which will be used several times (see Section 6.1 for technical assumptions).

- ⋆ In the strongly subcritical case (SS) ($\mathbb{E}(f'(1)\log(f'(1))) < 0$), there exist $c, \alpha_k > 0$ such that

$$\mathbb{P}_k(Z_n > 0) \sim c\alpha_k \mathbb{E}(f'(1))^n, \quad \alpha_1 = 1, \quad (n \to \infty). \tag{7.2}$$

- ⋆ In the intermediate subcritical case (IS) ($\mathbb{E}(f'(1)\log(f'(1))) = 0$), there exist $c, \alpha_k > 0$ such that

$$\mathbb{P}_k(Z_n > 0) \sim c\alpha_k n^{-1/2} \mathbb{E}(f'(1))^n, \quad \alpha_1 = 1, \quad (n \to \infty). \tag{7.3}$$

- ⋆ In the weakly subcritical case (WS) ($0 < \mathbb{E}(f'(1)\log(f'(1)))$), there exist $c, \alpha_k > 0$ such that

$$\mathbb{P}_k(Z_n > 0) \sim c\alpha_k n^{-3/2} \gamma^n, \quad \alpha_1 = 1, \quad (n \to \infty). \tag{7.4}$$

7.2 Subcriticality starting from several particles

We give here the asymptotic of survival probabilities starting with k particles. Then we determine how many initial particles survive conditionally on non extinction of particles and we characterize the sequence of environments which are selected by this conditioning. Finally we consider the Yaglom quasistationary distributions of $(Z_n)_{n\in\mathbb{N}}$ and the associated Q-process. In the (SS) case, results are those expected, i.e. they are analogous to those of a GW process. In the (IS) case, results are different for the Yaglom quasistationary distribution and the Q-process. In the (WS) case, all results are different.

We label by $i \in \mathbb{N}$ each particle of the initial population and denote by $Z_n^{(i)}$ the number of descendants of particle i at generation n.

Thus $(Z_n^{(i)})_{n\in\mathbb{N}}$ are identically distributed BPRE ($i \in \mathbb{N}$), with common distribution $(Z_n)_{n\in\mathbb{N}}$ starting with one particle. Conditionally on the environments, these processes are independent. In other words, for all $n, k, l_i \in \mathbb{N}$,

$$\mathbb{P}(Z_n^{(i)} = l_i,\ 1 \leq i \leq k \mid f_0, ..., f_{n-1}) = \Pi_{i=1}^k \mathbb{P}(Z_n^{(1)} = l_i \mid f_0, ..., f_{n-1}).$$

We denote by $\mathbb{P}_k$ the probability associated with k initial particles. Then, under $\mathbb{P}_k$, $(Z_n)_{n\in\mathbb{N}}$ is a.s. equal to

$$\big(\sum_{i=1}^k Z_n^{(i)}\big)_{n\in\mathbb{N}}.$$

7.2.1 Survival probabilities starting with several particles

Note that $x \mapsto \mathbb{E}\big(f'(1)^x \log(f'(1))\big)$ increases with x.

Proposition 7.2.1. *For every $k \in \mathbb{N}^*$,*

(i) If $\mathbb{E}\big(f'(1)^k \log(f'(1))\big) < 0$, then there exists $c_k > 0$ such that

$$\mathbb{P}(Z_n^{(1)} > 0,\ Z_n^{(2)} > 0,\ ...,\ Z_n^{(k)} > 0) \overset{n\to\infty}{\sim} c_k \mathbb{E}(f'(1)^k)^n$$

and $\mathbb{E}(f'(1)^k) < \mathbb{E}(f'(1)^{k-1}) < ... < \mathbb{E}(f'(1))$.

(ii) If $\mathbb{E}\big(f'(1)^k \log(f'(1))\big) = 0$, then there exists $c_k > 0$ such that

$$\mathbb{P}(Z_n^{(1)} > 0,\ Z_n^{(2)} > 0,\ ...,\ Z_n^{(k)} > 0) \overset{n\to\infty}{\sim} c_k n^{-1/2} \mathbb{E}(f'(1)^k)^n.$$

(iii) If $\mathbb{E}\big(f'(1)^k \log(f'(1))\big) > 0$, then there exists $c_k > 0$ such that

$$\mathbb{P}(Z_n^{(1)} > 0,\ Z_n^{(2)} > 0,\ ...,\ Z_n^{(k)} > 0) \overset{n\to\infty}{\sim} c_k n^{-3/2} \widetilde{\gamma}^n,$$

with $\widetilde{\gamma} = \inf_{s\in\mathbb{R}^+}\{\mathbb{E}(f'(1)^s)\} \in]0,1[$ and $c = c_1 \geq c_2 \geq ... \geq c_k$.

Moreover, in the (IS+WS) case, $\widetilde{\gamma} = \gamma$. In the (SS) case, $\widetilde{\gamma} < \gamma = \mathbb{E}(f'(1))$.

The proof is given in Section 7.3.1 and uses the case where the probability generating function f is a.s. linear fractional. Indeed in this case the survival probability in a given environment can then be computed explicitly since linear fractional generating functions are stable by composition (see Preliminaries Section).

In the (SS+IS) case, the asymptotic probability of survival of particles is proportional to the number of initial particles, as stated below. This is not surprising and well known for subcritical GW process. But this does not hold in the (WS) case. Recall that α_k is defined as $\lim_{n\to\infty} \mathbb{P}_k(Z_n > 0)/\mathbb{P}_1(Z_n > 0)$.

Theorem 7.2.2. *In the (SS+IS) case, for every $k \in \mathbb{N}$, $\alpha_k = k$.*

In the (WS) case, $\alpha_k \to \infty$ as $k \to \infty$ and there exists $M_+ > 0$ such that

$$\alpha_k \leq M_+ \log(k) k^{\alpha}, \quad (k \geq 2),$$

where $\alpha \in]0,1[$ is given by (7.1).
Assuming further $\mathbb{E}(f'(1)^{1/2} \log(f'(1))) > 0$ (i.e. $\alpha < 1/2$) and $f''(1)/f'(1)$ is bounded, there exists $M_- > 0$ such that

$$\alpha_k \geq M_- \log(k) k^{\alpha}, \quad (k \in \mathbb{N}).$$

One can naturally conjecture that the last result still holds for $1/2 \leq \alpha < 1$. The proof also uses the linear fractional case where, conditionally given the environments, the survival probability is related to a random walk whose jumps are the log of means of the reproduction law of the environments. That's why we first need to prove a result about random walk with negative drift conditioned to be larger than $-x < 0$ (see Appendix). One way to generalize the last result of the theorem above to the case $\mathbb{E}(f'(1)^{1/2} \log(f'(1))) > 0$ (i.e. $\alpha < 1/2$) would be to improve Lemma 7.4.1.

7.2.2 Survival of initial particles conditionally on non-extinction

We wonder now how many initial particles survive when we condition by the survival of the whole population of particles. We have the following elementary consequence of Proposition 7.2.1.

Proposition 7.2.3. *In the (SS+IS) case, for every $k \geq 1$,*

$$\lim_{n \to \infty} \mathbb{P}_k(\exists i \neq j,\ 1 \leq i, j \leq k,\ Z_n^{(i)} > 0,\ Z_n^{(j)} > 0 \ | Z_n > 0) = 0.$$

In the (WS) case, for every $k \geq 1$,

$$\lim_{n \to \infty} \mathbb{P}_k(\forall i,\ 1 \leq i \leq k,\ Z_n^{(i)} > 0 \ | Z_n > 0) > 0.$$

Thus, for (SS+IS) BPRE, conditionally on the survival of the population, only one initial particle survives, as for GW. But for (WS) BPRE, several initial particles survive with positive probability. Forthcoming Theorem 7.2.5 gives an interpretation of this property in terms of selection of favorable environments by conditioning on non-extinction. See Section 6.3 in [13] for an application of this result to a branching model for cell division with parasite infection, where we need to determine if several parasites survive in contaminated cells. In the same vein, see [38] for results on the reduced process associated with subcritical BPRE in the linear fractional case. In the (WS) case, the number of particles of the reduced process is not a.s. equal to 1 in the first generations.

What happens when the number of initial particles tends to infinity in the (WS) case ? As stated below, conditionally on non-extinction, the number of initial particles which survive is finite a.s. but not bounded, when the initial number of particles tend to infinity. More precisely, denote by N_n the number of particles in generation 0 whose descendance is alive in generation n. That is, starting with k initial particles :

$$N_n := \#\{1 \leq i \leq k \ : \ Z_n^{(i)} > 0\}.$$

Theorem 7.2.4. *In the (WS) case, assuming* $\mathbb{E}(f'(1)^{1/2}\log(f'(1))) > 0$ *(i.e.* $\alpha < 1/2$*) and* $f''(1)/f'(1)$ *is bounded, there exist* $A_l \downarrow_{l\to\infty} 0$ *such that for all* $k \geq l \geq 0$*,*

$$\limsup_{n\to\infty} \mathbb{P}_k(N_n \geq l \mid Z_n > 0) \leq A_l.$$

Moreover, for every $l \in \mathbb{N}^*$*,*

$$\liminf_{k\to\infty} \liminf_{n\to\infty} \mathbb{P}_k(N_n = l \mid Z_n > 0) > 0.$$

Thus, in the conditions of the theorem,

$$\limsup_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}_k(N_n \geq l \mid Z_n > 0) \leq A_l, \quad \text{with } A_l \downarrow_{l\to\infty} 0.$$

7.2.3 Selection of environments conditionally on non-extinction

We characterize here the sequence of environments which are selected by conditioning on the survival of particles.

We denote by $\mathcal{F}$ the set of generating functions and for every $\mathbf{g}_n = (g_0, \ldots, g_{n-1}) \in \mathcal{F}^n$, we denote by $Z_{\mathbf{g}_n}$ the value at generation n of the time inhomogeneous branching process whose reproduction law at generation $l \leq n-1$ has generating function g_l. Thus, for every $k \geq 1$,

$$\mathbb{E}_k(s^{Z_{\mathbf{g}_n}}) = g_0 \circ g_1 \circ \cdots \circ g_{n-1}(s)^k \qquad (0 \leq s \leq 1). \tag{7.5}$$

And we denote by $p(\mathbf{g}_n)$ the survival probability of a particle in environment $\mathbf{g}_n$. That is,

$$p(\mathbf{g}_n) := \mathbb{P}_1(Z_{\mathbf{g}_n} > 0). \tag{7.6}$$

Denote by $\mathbf{f}_n$ the sequence of environments until time n, i.e.

$$\mathbf{f}_n := (f_0, f_1, \ldots, f_{n-1}).$$

In the subcritical case, $p(\mathbf{f}_n) = 0$ a.s. since $(Z_n)_{n\in\mathbb{N}}$ becomes extinct a.s. Roughly speaking, the sequences of environments have a.s. zero survival probability. In the

(SS+IS) case, conditioning on the survival of particles does not change this fact, but it does in the (WS) case, as we can guess using Proposition 7.2.3. Actually, we prove that in the (WS) case, the sequence of environments which are selected by conditioning by $Z_n > 0$ have a.s. a positive survival probability. Thus, they are 'supercritical'. In [2], authors had already remarked this supercritical behavior of the BPRE $(Z_n)_{n\in\mathbb{N}}$ in the (WS) case by giving an analogy of the Kesten-Stigum theorem, i.e. the convergence of Z_n/m^n.

Theorem 7.2.5. *In the (SS+IS) case, for all $k \in \mathbb{N}^*$, $\epsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}_k(p(\mathbf{f}_n) \geq \epsilon \mid Z_n > 0) = 0.$$

In the (WS) case, for every $k \in \mathbb{N}^*$,

$$\liminf_{n\to\infty} \mathbb{P}_k(p(\mathbf{f}_n) \geq \epsilon \mid Z_n > 0) \stackrel{\epsilon\to 0+}{\longrightarrow} 1.$$

This supercritical behavior in the (WS) case disappears as k tends to infinity. That is, the survival probability of selected sequences of environments tends to 0 as the number of particles grows to infinity.

Proposition 7.2.6. *In the (WS) case, for every $\epsilon > 0$,*

$$\limsup_{n\to\infty} \mathbb{P}_k(p(\mathbf{f}_n) \geq \epsilon \mid Z_n > 0) \stackrel{k\to\infty}{\longrightarrow} 0.$$

In other words, conditionally on the survival of Z_n, the more initial particles there are, the less environments need to be favorable to allow the survival of particles, and the less likely it is for a given particle to survive. That's why, letting the number of initial particles tend to infinity does not make the number of survival initial particles tend to infinity, as stated in Theorem 7.2.4.

7.2.4 Yaglom quasistationary distributions

We focus now on the Yaglom quasistationary distribution of $(Z_n)_{n\in\mathbb{N}}$ (see Preliminaries for existence and references). For the GW process, this distribution does not depend on the initial number of particles and is characterized by a functional equation. This result still holds for (SS) BPRE. Indeed, starting with several particles, conditionally on the survival of one given particle, the others become extinct (see Proposition 7.2.3).

Theorem 7.2.7. *In the (SS+IS) case, for every $k \geq 1$, the BPRE Z_n starting from k and conditioned to be nonzero converges in distribution as $n \to \infty$ to a r.v. Υ which does not depend on k. Moreover, the generating function G of Υ verifies*

$$\mathbb{E}(G(f(s))) = \mathbb{E}(f'(1))G(s) + 1 - \mathbb{E}(f'(1)), \qquad G(0) = 0.$$

In the (SS) case, G is the unique generating function which satisfies the functional equation above and $G'(1) < \infty$.

In the (WS) case, for every $k \geq 1$, the BPRE Z_n starting from k and conditioned to be nonzero converges in distribution as $n \to \infty$ to a r.v. Υ_k, whose generating function G_k verifies

$$\mathbb{E}(G_k(f(s))) = \gamma G_k(s) + 1 - \gamma, \qquad G_k(0) = 0.$$

In the (WS) case, an open question is to determine if the quasistationary distribution Υ_k depends on the initial number k of particles. We know that for every $k \geq 1$, G_k verifies the same functional equation given above but we do not know if the solution is unique.
Moreover, we can prove the equality of the quasistationary distributions starting with two different numbers of particles in the following case. If $Z_1 \in \{0, 1, N\}$ for some $N \in \mathbb{N}^*$, then $\Upsilon_1 \stackrel{d}{=} \Upsilon_N$. (see Section 7.2.4 for the proof). Other observations also lead us to believe that quasistationary distributions Υ_k might not depend on k.

7.2.5 Q-process associated with a BPRE

The Q-process $(Y_n)_{n\in\mathbb{N}}$ starting from k particles associated to the BPRE $(Z_n)_{n\in\mathbb{N}}$ is defined for all $l_1, l_2, ..., l_n \in \mathbb{N}$, by

$$\mathbb{P}_k(Y_1 = l_1, ..., Y_n = l_n) = \lim_{p\to\infty} \mathbb{P}_k(Z_1 = l_1, ..., Z_n = l_n | Z_{n+p} > 0).$$

This is the BPRE $(Z_n)_{n\in\mathbb{N}}$ conditioned to survive in the distant future. See [8] for details in the case of GW processes. In the (SS) case, the Q-process converges in distribution to the size biased Yaglom distribution, as for GW process. Finer results have been obtained in [1]. In the (IS+WS) case, the Q-process is transient. That is, the population needs to grow largely in the first generations so that it can survive.

Proposition 7.2.8. $\star$ *In the (SS) case, for every $k \in \mathbb{N}^*$, for all $l_1, l_2, ..., l_n \in \mathbb{N}$,*

$$\mathbb{P}_k(Y_1 = l_1, ..., Y_n = l_n) = \mathbb{E}(f'(1))^{-n} \frac{l_n}{k} \mathbb{P}_k(Z_1 = l_1, ..., Z_n = l_n).$$

Moreover $(Y_n)_{n\in\mathbb{N}}$ converges in distribution to the size biased Yaglom distribution.

$$\forall\ l \geq 0, \qquad \mathbb{P}_k(Y_n = l) \stackrel{n\to\infty}{\longrightarrow} \frac{l\mathbb{P}(\Upsilon = l)}{\mathbb{E}(\Upsilon)}.$$

$\star$ *In the (IS) case, for every $k \in \mathbb{N}^*$, for all $l_1, l_2, ..., l_n \in \mathbb{N}$,*

$$\mathbb{P}_k(Y_1 = l_1, ..., Y_n = l_n) = \mathbb{E}(f'(1))^{-n} \frac{l_n}{k} \mathbb{P}_k(Z_1 = l_1, ..., Z_n = l_n).$$

Moreover $Y_n \to \infty$ in probability as $n \to \infty$.

$\star$ *In the (WS) case, for every $k \in \mathbb{N}^*$, for all $l_1, l_2, ..., l_n \in \mathbb{N}$,*

$$\mathbb{P}_k(Y_1 = l_1, ..., Y_n = l_n) = \gamma^{-n} \frac{\alpha_{l_n}}{\alpha_k} \mathbb{P}_k(Z_1 = l_1, ..., Z_n = l_n).$$

Moreover Y_n tends to infinity a.s.

We focus now on the environments of the Q-process. We endow $\mathcal{F}$ with distance d given by the infinity norm

$$d(f,g) = \| f - g \|_\infty$$

and we denote by $\mathbf{B}_0(\mathcal{F})$ the Borel σ-field.

We introduce the probability ν_k on $(\mathcal{F}^{\mathbb{N}}, \mathbf{B}_0(\mathcal{F})^{\otimes \mathbb{N}})$ which gives the distribution of the environments when the BPRE $(Z_n)_{n \in \mathbb{N}}$ starting from k particles is conditioned to survive. Using Kolomogorov Theorem, it can be specified by its projection on $(\mathcal{F}^p, \mathbf{B}_0(\mathcal{F})^{\otimes p})$ for every $p \in \mathbb{N}$, denoted by $\nu_{k \ |\mathcal{F}^p}$,

$$\begin{aligned} \nu_{k \ |\mathcal{F}^p}(\mathrm{d}\mathbf{g}_p) &:= \lim_{n \to \infty} \mathbb{P}_k\big(\mathbf{f}_p \in \mathrm{d}\mathbf{g}_p | Z_{n+p} > 0\big) \\ &= \gamma^{-p} \mathbb{P}\big(\mathbf{f}_p \in \mathrm{d}\mathbf{g}_p\big) \sum_{l=1}^{\infty} \mathbb{P}_k(Z_{\mathbf{g}_p} = l) \frac{\alpha_l}{\alpha_k}, \end{aligned} \tag{7.7}$$

with $\mathbf{f}_p = (f_0, \ldots, f_{p-1})$ and $\gamma = \mathbb{E}(f'(1))$ in the (SS+IS) case. The limit is the weak limit of probabilities on $(\mathcal{F}^p, \mathbf{B}_0(\mathcal{F})^{\otimes p})$ (see [24] for definition and Section 7.3.5 for the proof), which we endow with the distance d_p given by

$$d_p\big((g_0, \ldots, g_{p-1}), (h_0, \ldots, h_{p-1})\big) = \sup\{\| g_i - h_i \|_\infty \colon 0 \le i \le p-1\}. \tag{7.8}$$

For every $\mathbf{g} \in \mathcal{F}^{\mathbb{N}}$, we denote by $\mathbf{g}|n$ the first n coordinates of $\mathbf{g} \in \mathcal{F}^{\mathbb{N}}$ and we introduce the survival probability in environment $\mathbf{g} \in \mathcal{F}^{\mathbb{N}}$:

$$p(\mathbf{g}) = \lim_{n \to \infty} \downarrow \mathbb{P}(Z_{\mathbf{g}|n} > 0).$$

One can naturally conjecture an analogous of Theorem 7.2.5 and Proposition 7.2.6. That is, for every $k \in \mathbb{N}^*$,

In the (SS+IS) case, $\quad \nu_k(\{\mathbf{g} \in \mathcal{F}^{\mathbb{N}} : p(\mathbf{g}) = 0\}) = 1.$

In the (WS) case, $\quad \nu_k(\{\mathbf{g} \in \mathcal{F}^{\mathbb{N}} : p(\mathbf{g}) > 0\}) = 1 \quad$ and $\quad \nu_k(p(\mathbf{f}) \in \mathrm{d}x) \overset{n \to \infty}{\Longrightarrow} \delta_0(\mathrm{d}x).$

A perspective is to characterize the tree of particles when we condition by the survival of particles, i.e. the tree of particles of the Q-process. Informally, for GW process, this gives a spine with finite iid subtrees (see [42, 69]). This fact still holds in the (SS+IS) case but we will observe a 'multispine tree' in the (WS) case.

7.3 Proofs

First we give the main Notations and results for the proofs. We use the particular case when generating functions are a.s. linear fractional. In that case, the survival probability for a given environment is a functional of the random walk which sums the *log* of the successive means of environments (see (6.6)). Using results the random walk with negative drift proved in Appendix (Section 7.4), this enables us to control the survival probability conditionally on the environments, in the linear fractional case and then for general BPRE (Lemma 7.3.1 below). Using that conditionally on the sequence of environments, particles are independent, we get survival probabilities starting with several particles and then integrate with respect to environments.

Set for every $n \in \mathbb{N}$,

$$X_n := \log(f_n'(1)), \qquad S_n = \sum_{i=0}^{n-1} X_i \quad (S_0 = 0),$$

and

$$L_n = \min\{S_i \ : \ 1 \leq i \leq n\}.$$

Recall that $\mathcal{F}$ is the set of generating functions and

$$\mathbf{f}_n = (f_0, f_1, ..., f_{n-1}).$$

For every $\mathbf{g}_n = (g_0, \dots, g_{n-1}) \in \mathcal{F}^n$, we denote by $Z_{\mathbf{g}_n}$ the value at generation p of the time inhomogeneous branching process whose reproduction law at generation l has generating function g_l. And we denote by $p(\mathbf{g}_n)$ the survival probability of a particle in environment $\mathbf{g}_n$. That is,

$$p(\mathbf{g}_n) := \mathbb{P}_1(Z_{\mathbf{g}_n} > 0) = \mathbb{P}_1(Z_n > 0 \mid \mathbf{f}_n = \mathbf{g}_n). \tag{7.9}$$

Roughly speaking, we prove now that

$$p(\mathbf{f}_n) \approx e^{L_n} \quad \text{a.s..}$$

Lemma 7.3.1. *For every $n \in \mathbb{N}$, we have*

$$p(\mathbf{f}_n) \leq e^{L_n} \quad a.s.$$

Moreover if $\mathbb{E}(f'(1)^{1/2}\log(f'(1))) > 0$ (i.e. $0 < \alpha < 1/2$) and $f''(1)/f'(1)$ is bounded, then there exists $\mu \geq 1$ such that for all $n \in \mathbb{N}$ and $x \in]0, 1]$,

$$\mathbb{P}(p(\mathbf{f}_n) \geq x) \geq \mathbb{P}(L_n \geq \log(\mu x))/4.$$

Proof. For the upper bound, note that for every $\mathbf{g}_n \in \mathcal{F}^n$, we have,

$$p(\mathbf{g}_n) = \mathbb{P}_1(Z_{\mathbf{g}_n} > 0) \leq \mathbb{E}_1(Z_{\mathbf{g}_n}) = \Pi_{i=0}^{n-1} g_i'(1).$$

Thus $p(\mathbf{f}_n) \leq e^{S_n}$ a.s. As is usual, adding that $p(\mathbf{f}_n)$ decreases a.s. ensures that

$$p(\mathbf{f}_n) \leq e^{L_n} \quad a.s.$$

For the lower bound, recall that

$$P_i := f_{n-i}'(1)...f_{n-1}'(1), \qquad (P_0 = 1),$$

and use (6.9) and (6.6) to get

$$p(\mathbf{f}_n) = \mathbb{P}(Z_n > 0 \mid \mathbf{f}_n) \geq \frac{\widetilde{P}_n}{1 + \sum_{i=0}^{n-1} \frac{\widetilde{f}''_{n-i-1}(1)}{2\widetilde{f}'_{n-i-1}(1)} \widetilde{P}_i} = \frac{P_n}{1 + \sum_{i=0}^{n-1} \frac{f''_{n-i-1}(1)}{f'_{n-i-1}(1)} P_i} \quad \text{a.s..}$$

We assume now that $C = \big(1 + \text{ess sup}(\frac{f''(1)}{f'(1)})\big)^{-1} > 0$. Denote by

$$S_i' := \log(f_{n-i}'(1)) + ... + \log(f_{n-1}'(1)) \quad i \geq 1, \qquad S_0' = 0,$$

so that $P_i = \exp(S_i')$. We then have

$$p(\mathbf{f}_n) \geq C \frac{e^{S_n'}}{2\sum_{i=0}^{n-1} e^{S_i'}} \geq \frac{C}{2} \frac{e^{S_n' - \max\{S_j' : 0 \leq j \leq n\}}}{\sum_{i=0}^{n} e^{S_i' - \max\{S_j' : 0 \leq j \leq n\}}} \quad \text{a.s.}$$

Thus,

$$p(\mathbf{f}_n) \geq \frac{C}{2} \frac{e^{L_n}}{\sum_{i=0}^{n} e^{L_n - S_i}}. \tag{7.10}$$

As $\alpha < 1/2$, forthcoming Corollary 7.4.2 in Appendix (Section 7.4) ensures that there exists $\beta > 0$ such that for all $n \in \mathbb{N}$ and $x > 0$,

$$\begin{aligned} \mathbb{P}(p(\mathbf{f}_n) \geq x) &\geq \mathbb{P}(L_n \geq \log(2\beta x/C)) \mathbb{P}\big(\sum_{i=0}^{n} e^{L_n - S_i} \leq \beta \mid L_n \geq \log(2\beta x/C)\big) \\ &\geq \mathbb{P}(L_n \geq \log(\mu x))/4, \end{aligned}$$

writing $\mu = \min(1, 2\beta/C)$. □

Moreover, using independence of particles conditionally on environments, we have

$$\mathbb{P}_k(Z_n > 0) = \int_{\mathcal{F}^n} \mathbb{P}_k(Z_{\mathbf{g}_n} > 0) \mathbb{P}(\mathbf{f}_n \in \mathrm{d}\mathbf{g}_n)$$

$$\begin{aligned} &= \int_{\mathcal{F}^n} (1-(1-\mathbb{P}_1(Z_{\mathbf{g}_n}>0))^k)\mathbb{P}(\mathbf{f}_n \in \mathrm{d}\mathbf{g}_n) \\ &= \int_{\mathcal{F}^n} \mathbb{P}(p(\mathbf{f}_n)\in \mathrm{d}x)(1-(1-x)^k), \end{aligned} \tag{7.11}$$

and

$$\begin{aligned} \alpha_k &= \lim_{n\to\infty} \mathbb{P}_k(Z_n>0)/\mathbb{P}_1(Z_n>0) \\ &= \lim_{n\to\infty} \int_0^1 (1-(1-x)^k)\frac{\mathbb{P}(p(\mathbf{f}_n)\in \mathrm{d}x)}{\mathbb{P}_1(Z_n>0)}. \end{aligned} \tag{7.12}$$

7.3.1 Proofs of Section 7.2.1

We split the proof of Proposition 7.2.1 into three parts.

Proof of Proposition 7.2.1 (i). We follow the proof of Theorem 1.2 (a) in [46] and introduce the probability $\widetilde{\mathbb{P}}$ such that under $\widetilde{\mathbb{P}}$, the environments still are iid and their law is given by

$$\widetilde{\mathbb{P}}(f\in \mathrm{d}g) = \mathbb{E}(f'(1)^k)^{-1} g'(1)^k \mathbb{P}(f\in \mathrm{d}g).$$

Then, writing $P_n = f_0'(1)...f_{n-1}'(1)$ $(P_0=1)$, we have

$$\mathbb{P}(Z_n^{(1)}>0,...,Z_n^{(k)}>0) = \mathbb{E}((1-F_n(0))^k) = \mathbb{E}(f'(1)^k)^n \widetilde{\mathbb{E}}((1-F_n(0))/P_n)^k).$$

As $\mathbb{E}(f'(1)^k \log(f'(1)))<0$, then $\widetilde{\mathbb{E}}(\log(f'(1)))<0$ and Theorem 5 in [6] ensures that

$$C = \lim_{n\to\infty} \frac{1-F_n(0)}{P_n}$$

exists $\widetilde{\mathbb{P}}$ a.s. and belongs to $]0,1]$. Thus, as $n\to\infty$,

$$\mathbb{P}(Z_n^{(1)}>0,...,Z_n^{(k)}>0) \sim \mathbb{E}(f'(1)^k)^n \widetilde{\mathbb{E}}(C^k).$$

Add that $s\mapsto \mathbb{E}(f'(1)^s)$ decreases for $s\in[0,\alpha]$ and $k<\alpha$ to complete the proof. □

Proof of Proposition 7.2.1 (iii). We follow the proof of Theorem 1.2 (c) in [46].

STEP 1. First we consider the linear fractional case and use results of the previous chapter (see (6.5)). In that case, by (6.7),

$$\mathbb{P}(Z_n^{(1)} > 0, ..., Z_n^{(k)} > 0 |\ f_0, ...,\ f_{n-1}) = \left(1 + \sum_{i=0}^{n-1} \frac{f''_{n-i-1}(1)}{2f'_{n-i-1}(1)} P_i\right)^{-k} P_n^k.$$

Define $\widetilde{\gamma}$ by

$$\widetilde{\gamma} = \inf_{s\in\mathbb{R}^+} \{\mathbb{E}(f'(1)^s)\} = \mathbb{E}(f'(1)^{\widetilde{\alpha}}),$$

where $0 < \widetilde{\alpha} < k$ since $\mathbb{E}(f'(1)^k \log(f'(1))) > 0$. Let $\mathbb{P}_{\widetilde{\alpha}}$ be the probability given by

$$\mathbb{P}_{\widetilde{\alpha}}(f \in \mathrm{d}g) = \widetilde{\gamma}^{-1} g'(1)^{\widetilde{\alpha}} \mathbb{P}(f \in \mathrm{d}g).$$

Then

$$\mathbb{P}(Z_n^{(1)} > 0, ..., Z_n^{(k)} > 0) = \widetilde{\gamma}^n \mathbb{E}_{\widetilde{\alpha}}\left[\left(1 + \sum_{i=0}^{n-1} \frac{f''_i(1)}{2f'_i(1)} P_i\right)^{-k} P_n^{k-\widetilde{\alpha}}\right].$$

As $\mathbb{E}_{\widetilde{\alpha}}(\log(f'(1))) = 0$, we apply Theorem 2.1 in [46] with

$$\phi(x) = x^{k-\widetilde{\alpha}}, \quad \psi(x) = (1+x)^{-k}, \quad 0 < k - \widetilde{\alpha} < k,$$

so there exists $c_k > 0$ such that, as $n \to \infty$,

$$\mathbb{P}(Z_n^{(1)} > 0, ..., Z_n^{(k)} > 0) \sim c_k \widetilde{\gamma}^n n^{-3/2}.$$

STEP 2. For the general case, we can use STEP 1. Indeed, by (6.9), there exists a BPRE $(\widetilde{Z}_n)_{n\in\mathbb{N}}$ such that $\widetilde{f}$ is a.s. linear fractional, $\widetilde{f}'(1) = f'(1)$ and

$$\mathbb{P}(Z_n^{(1)} > 0,\ Z_n^{(2)} > 0,\ ...,\ Z_n^{(k)} > 0) \geq \mathbb{P}(\widetilde{Z}_n^{(1)} > 0,\ \widetilde{Z}_n^{(2)} > 0,\ ...,\ \widetilde{Z}_n^{(k)} > 0).$$

By STEP 1, this leads to the existence of $c_k(1) > 0$ such that

$$\mathbb{P}(Z_n^{(1)} > 0,\ Z_n^{(2)} > 0,\ ...,\ Z_n^{(k)} > 0) \geq c_k(1)\gamma^n n^{-3/2}. \tag{7.13}$$

Note that by inclusion-exclusion principle, we have

$$\mathbb{P}_k(Z_n > 0) = \sum_{i=1}^{k} (-1)^{i+1} \binom{k}{i} \mathbb{P}(Z_n^{(1)} > 0, ..., Z_n^{(i)} > 0). \tag{7.14}$$

Moreover, (7.4) ensure the convergence of

$$\gamma^{-n} n^{3/2} \mathbb{P}_1(Z_n > 0)$$

to $c\alpha_1$. By induction, it gives the convergence of

$$\gamma^{-n} n^{3/2} \mathbb{P}(Z_n^{(1)} > 0,\ Z_n^{(2)} > 0,\ ...,\ Z_n^{(k)} > 0).$$

to a constant c_k, which is positive by (7.13).

To complete the proof note that $\gamma = \widetilde{\gamma}$ iff $\mathbb{E}(f'(1)^s)]'(1) \geq 0$, i.e. in the (IS+WS) case. $\square$

Proof of Proposition 7.2.1 (ii). The proof is close to the previous one. First, we consider the linear fractional case and Introducing again the probability $\widetilde{\mathbb{P}}$ defined by

$$\widetilde{\mathbb{P}}(f \in \mathrm{d}g) = \mathbb{E}(f'(1)^k)^{-1} g'(1)^k \mathbb{P}(f \in \mathrm{d}g).$$

Using again (6.7), we get then

$$\mathbb{P}(Z_n^{(1)} > 0, ..., Z_n^{(k)} > 0) = \mathbb{E}(f'(1)^k)^n \widetilde{\mathbb{E}}\left[\left(1 + \sum_{i=0}^{n-1} \frac{f''_{n-i-1}(1)}{2f'_{i-i-1}(1)} P_i\right)^{-k}\right].$$

As $\widetilde{\mathbb{E}}(\log(f'(1)) = 0$, we can use again Theorem 2.1 in [46] and conclude in the linear fractional case.

The general case can be proved following STEP 2 in the previous proof. □

Proof of Theorem 7.2.2. Computation of α_k in the (SS+IS) case. In the (SS+IS) case, Proposition 7.2.3 and (7.14) ensure that for every $k \in \mathbb{N}$,

$$\mathbb{P}_k(Z_n > 0) \sim k\mathbb{P}(Z_n > 0), \qquad (n \to \infty).$$

Then,

$$\alpha_k = k.$$

This gives the first result.

Limit of α_k in the (WS) case. Note that $\mathbb{P}_1(Z_{p+n} > 0) = \sum_{k=1}^{\infty} \mathbb{P}_1(Z_p = k)\mathbb{P}_k(Z_n > 0)$. Then,

$$\frac{\mathbb{P}_1(Z_{p+n} > 0)}{\mathbb{P}_1(Z_n > 0)} = \sum_{k=1}^{\infty} \mathbb{P}_1(Z_p = k) \frac{\mathbb{P}_k(Z_n > 0)}{\mathbb{P}_1(Z_n > 0)}. \tag{7.15}$$

First, $\mathbb{P}(\cup_{i=1}^k \{Z_n^{(i)} > 0\}) \leq \sum_{i=1}^k \mathbb{P}(Z_n^{(i)} > 0)$, which gives

$$\mathbb{P}_k(Z_n > 0)/\mathbb{P}_1(Z_n > 0) \leq k.$$

Moreover $\sum_{k=1}^{\infty} \mathbb{P}_1(Z_p = k)k = \mathbb{E}(Z_p) < \infty$ and

$$\mathbb{P}_k(Z_n > 0)/\mathbb{P}_1(Z_n > 0) \overset{n\to\infty}{\longrightarrow} \alpha_k,$$

then, by bounded convergence, we get

$$\sum_{k=1}^{\infty} \mathbb{P}_1(Z_p = k) \frac{\mathbb{P}_k(Z_n > 0)}{\mathbb{P}_1(Z_n > 0)} \stackrel{n\to\infty}{\longrightarrow} \sum_{k=1}^{\infty} \mathbb{P}_1(Z_p = k)\alpha_k.$$

Then, using again (7.4), letting $n \to \infty$ in (7.15) yields

$$\gamma^p = \sum_{k=1}^{\infty} \mathbb{P}_1(Z_p = k)\alpha_k.$$

Assuming that $(\alpha_k)_{k\in\mathbb{N}}$ is bounded by A leads to

$$\gamma^p \leq A\mathbb{P}_1(Z_p > 0).$$

Letting $p \to \infty$ leads to a contradiction with (7.4). Adding that α_k increases ensures that $\alpha_k \to \infty$ as $k \to \infty$.

Upper bound of α_k in the (WS) case. Recall (7.12),

$$\alpha_k = \lim_{n\to\infty} \int_0^1 (1-(1-x)^k) \frac{\mathbb{P}(p(\mathbf{f}_n) \in \mathrm{d}x)}{\mathbb{P}_1(Z_n > 0)}.$$

Using the first inequality of Lemma 7.3.1 and $x \mapsto 1-(1-x)^k$ grows with x on $[0,1]$, we have

$$\alpha_k \leq \limsup_{n\to\infty} \int_0^1 (1-(1-x)^k) \frac{\mathbb{P}(\exp(L_n) \in \mathrm{d}x)}{\mathbb{P}_1(Z_n > 0)}.$$

By (7.21), we can use Fatou's Lemma and (7.22) gives

$$\alpha_k \leq \int_0^1 (1-(1-x)^k)\nu_+(\mathrm{d}x) \limsup_{n\to\infty} \frac{n^{-3/2}\gamma^n}{\mathbb{P}_1(Z_n > 0)}.$$

Thus, by (7.4) and definition of ν_+,

$$\alpha_k \leq c^{-1}c_+[1 + \int_0^1 (1-(1-x)^k)\log(1/x)x^{-\alpha-1}\mathrm{d}x].$$

Finally, using $1-(1-x)^k \leq kx$ and integration by parts,

$$\begin{aligned} & \int_0^1 (1-(1-x)^k)\log(1/x)x^{-\alpha-1}\mathrm{d}x \\ = \; & \int_0^{1/k} (1-(1-x)^k)\log(1/x)x^{-\alpha-1}\mathrm{d}x + \int_{1/k}^1 (1-(1-x)^k)\log(1/x)x^{-\alpha-1}\mathrm{d}x \\ \leq \; & k\int_0^{1/k} \log(1/x)x^{-\alpha}\mathrm{d}x + \log(k)\int_{1/k}^1 x^{-\alpha-1}\mathrm{d}x \end{aligned}$$

$$
\begin{aligned}
&\leq \; k(1-\alpha)^{-1}\big[\log(k)k^{\alpha-1} + \int_0^{1/k} x^{-\alpha}\mathrm{d}x\big] + \alpha^{-1}\log(k)(k^{\alpha}-1) \\
&\leq \; M(1+\log(k))k^{\alpha},
\end{aligned}
$$

for some $M > 0$. This ensures that $\alpha_k \leq c^{-1}c_+[1 + M(1+\log(k))k^{\alpha}]$ and ends the proof.

Lower bound of α_k in the (WS) case assuming further $\mathbb{E}(f'^{1/2}(1)\log(f'(1))) > 0$ (i.e. $\alpha < 1/2$) and $f''(1)/f'(1)$ is bounded.

By (7.4), Lemma 7.3.1 and (7.22), for every $x > 0$,

$$
\begin{aligned}
\liminf_{n\to\infty} \frac{\mathbb{P}(p(\mathbf{f}_n) \geq x)}{\mathbb{P}_1(Z_n > 0)} &= \liminf_{n\to\infty} \frac{\gamma^n n^{-3/2}}{\mathbb{P}_1(Z_n > 0)} \frac{\mathbb{P}(p(\mathbf{f}_n) \geq x)}{\gamma^n n^{-3/2}} \\
&\geq c^{-1}\frac{\mathbb{P}(L_n \geq \log(\mu x))}{\gamma^n n^{-3/2}} \\
&\geq (4c)^{-1}\nu_-([\mu x, 1]).
\end{aligned}
$$

Using (7.12), Fatou's Lemma ensures that,

$$
\begin{aligned}
\alpha_k &\geq (4c)^{-1}\int_0^{\mu^{-1}} (1-(1-x)^k)\nu_-(\mathrm{d}(\mu x)) \\
&\geq (4c)^{-1}c_-\mu^{\alpha}\int_0^{1/k} (1-(1-x)^k)[\log(1/x) - \log(\mu)]x^{-\alpha-1}\mathrm{d}x.
\end{aligned}
$$

For all $k \geq \mu^2$ and $x \in]0, 1/k]$, $\log(1/x) \geq 2\log(\mu)$. So for every $k \geq \mu^2$,

$$
\begin{aligned}
\alpha_k &\geq 2^{-1}\int_0^{1/k} (1-(1-x)^k)\log(1/x)x^{-\alpha-1}\mathrm{d}x \\
&\geq 2^{-1}k \inf_{x\in]0,1/k]}\left\{\frac{1-(1-x)^k}{kx}\right\}\int_0^{1/k} \log(1/x)x^{-\alpha}\mathrm{d}x \\
&\geq 2^{-1}k(1-(1-1/k)^k)\int_0^{1/k} \log(1/x)x^{-\alpha}\mathrm{d}x \\
&\geq 2^{-1}k(1-(1-1/k)^k)\log(k)\int_0^{1/k} x^{-\alpha}\mathrm{d}x \\
&\geq 2^{-1}\log(k)k^{\alpha}/(1-\alpha).
\end{aligned}
$$

This completes the proof. □

7.3.2 Proofs of Section 7.2.2

Proof of Proposition 7.2.3. The first part (i.e. the (SS+IS) case) follows from

$$\mathbb{P}_k(\exists i \neq j,\ 1 \leq i,j \leq k,\ Z_n^{(i)} > 0,\ Z_n^{(j)} > 0\,|Z_n > 0) \leq \binom{k}{2} \frac{\mathbb{P}(Z_n^{(1)} > 0,\ Z_n^{(2)} > 0)}{\mathbb{P}_k(Z_n > 0)},$$

asymptotics given by Proposition 7.2.1 (i-ii-iii) and equations (6.2) and (7.3).

The second part (i.e. the (WS) case) is directly derived from Proposition 7.2.1 (iii) and (7.4). □

Proof of Theorem 7.2.4. Denote by $N(\mathbf{g}_n)$ the number of initial particles which survive until generation n where successive reproduction laws are given by $\mathbf{g}_n$ (i.e. conditionally on $\mathbf{f}_n = \mathbf{g}_n$). Then, for all $1 \leq l \leq k$,

$$\begin{aligned}
\mathbb{P}_k(N_n = l) &= \int_{\mathcal{F}^n} \mathbb{P}(\mathbf{f}_n \in \mathrm{d}\mathbf{g}_n)\mathbb{P}_k(N(\mathbf{g}_n) = l) \\
&= \int_0^1 \mathbb{P}(p(\mathbf{f}_n) \in \mathrm{d}x)\binom{k}{l} x^l(1-x)^{k-l}.
\end{aligned}$$

Note that $x \in [0,1] \mapsto x^l(1-x)^{k-l}$ is positive, increases on $[0, l/k]$ and decreases on $[l/k, 1]$.

First, we prove the upper bound. Recalling Lemma 7.3.1,

$$p(\mathbf{f}_n) \leq \exp(L_n),$$

we get

$$\begin{aligned}
&\int_0^1 \mathbb{P}(p(\mathbf{f}_n) \in \mathrm{d}x)x^l(1-x)^{k-l} \\
= &\int_0^1 \mathbb{P}(p(\mathbf{f}_n) \in \mathrm{d}x,\ \exp(L_n) \leq l/k)x^l(1-x)^{k-l} \\
&\quad + \int_0^1 \mathbb{P}(p(\mathbf{f}_n) \in \mathrm{d}x,\ \exp(L_n) > l/k)x^l(1-x)^{k-l} \\
\leq &\int_0^1 \mathbb{P}(\exp(L_n) \in \mathrm{d}x)x^l(1-x)^{k-l} + \mathbb{P}(\exp(L_n) \in]l/k, 1])(l/k)^l(1-l/k)^{k-l}.
\end{aligned}$$

Moreover, by (7.20),

$$\limsup_{n\to\infty} \frac{\mathbb{P}(\exp(L_n) \in]l/k, 1])}{\gamma^n n^{-3/2}} \leq u(\log(k/l))(k/l)^{\alpha}.$$

Second, using again the variations of $x \in [0,1] \mapsto x^l(1-x)^{k-l}$ and (7.21), we get

$$\begin{aligned}
& \lim_{n\to\infty} \int_0^1 \frac{\mathbb{P}(\exp(L_n) \in \mathrm{d}x)}{n^{-3/2}\gamma^n} x^l(1-x)^{k-l} \\
\leq\ & \int_0^{l/k} \nu_+(\mathrm{d}x)x^l(1-x)^{k-l} + \nu_+([l/k,1])(l/k)^l(1-l/k)^{k-l} \\
\leq\ & c_+ \int_0^1 \log(1/x)x^{-\alpha-1}x^l(1-x)^{k-l}\mathrm{d}x \\
& \quad + c_+(1 + \int_{l/k}^1 \log(1/x)x^{-\alpha-1}\mathrm{d}x)(l/k)^l(1-l/k)^{k-l} \\
\leq\ & c_+ \int_0^1 \log(1/x)x^{-\alpha-1}x^l(1-x)^{k-l}\mathrm{d}x \\
& \quad + c_+\big(1 + \log(k/l)\frac{(k/l)^\alpha - 1}{\alpha}\big)(l/k)^l(1-l/k)^{k-l}.
\end{aligned}$$

Putting the three last inequalities together and using $u(\log(k/l)) \leq C(1 + \log(k/l))$ for some $C > 0$ ensures that there exists $D > 0$ such that

$$\begin{aligned}
& \limsup_{n\to\infty} \int_0^1 \frac{\mathbb{P}(p(\mathbf{f}_n) \in \mathrm{d}x)}{n^{-3/2}\gamma^n} x^l(1-x)^{k-l} \\
\leq\ & c_+ \int_0^1 \log(1/x)x^{-\alpha-1}x^l(1-x)^{k-l}\mathrm{d}x + D(1 + \log(k/l)(k/l)^\alpha)(l/k)^l(1-l/k)^{k-l}.
\end{aligned}$$

Moreover, denoting by B is the Beta function, we have

$$\begin{aligned}
& \int_0^1 \log(x)x^{-\alpha-1}x^l(1-x)^{k-l}\mathrm{d}x \\
=\ & \int_0^{1/k} \log(1/x)x^{l-\alpha-1}(1-x)^{k-l}\mathrm{d}x + \int_{1/k}^1 \log(1/x)x^{l-\alpha-1}(1-x)^{k-l}\mathrm{d}x \\
\leq\ & \int_0^{1/k} \log(1/x)x^{l-\alpha-1}\mathrm{d}x + \log(k)\int_{1/k}^1 x^{l-\alpha-1}(1-x)^{k-l}\mathrm{d}x \\
\leq\ & (l-\alpha)^{-1}\big[\log(k)k^{\alpha-l} + (l-\alpha)^{-1}k^{\alpha-l}\big] + \log(k)B(l-\alpha, k-l+1),
\end{aligned}$$

by integration by parts. By Stirling formula, there exists $C > 0$, and then $C', C'' > 0$ such that for all $1 \leq l \leq k$,

$$\begin{aligned}
\binom{k}{l} k^{-\alpha}B(l-\alpha, k-l+1) \ &\leq\ C\frac{k^{k-\alpha+1/2}}{l^{l+1/2}(k-l)^{k-l+1/2}} \frac{(l-\alpha)^{l-\alpha-1/2}(k-l+1)^{k-l+1/2}}{(k-\alpha+1)^{k-\alpha+1/2}} \\
&\leq\ C'\frac{(l-\alpha)^{l-\alpha-1/2}(k-l+1)^{k-l+1/2}}{l^{l+1/2}(k-l)^{k-l+1/2}} \\
&\leq\ C''\frac{1}{l^{1+\alpha}}, \qquad (7.16)
\end{aligned}$$

where the last inequality comes from the fact that $(1/x+1/2)\log(1+x)$ is bounded for $x \in [0,1]$, so that $(k-l+1/2)\log(1+1/(k-l))$ is bounded for $1 \leq l < k$.

Then, combining the three last inequalities gives

$$\begin{aligned}
\limsup_{n\to\infty} \frac{\mathbb{P}_k(N_n = l)}{k^\alpha \log(k) n^{-3/2}\gamma^n} &= \limsup_{n\to\infty} \frac{\binom{k}{l}}{k^\alpha \log(k)} \int_0^1 \frac{\mathbb{P}(\exp(L_n) \in \mathrm{d}x)}{n^{-3/2}\gamma^n} x^l (1-x)^{k-l} \\
&\leq (l-\alpha)^{-1}\big[\binom{k}{l} k^{-l} + (l-\alpha)^{-1} k^{-l}/\log(k) + C'' \frac{1}{l^{1+\alpha}}\big] \\
&\quad + D\binom{k}{l}(\log(k)k^{-\alpha} + l^{-\alpha})(l/k)^l (1-l/k)^{k-l}.
\end{aligned}$$

Again Stirling formula ensures that there exists $C''' > 0$ such that

$$\binom{k}{l} k^{-l} \leq C''' \frac{k^{k-l}}{(k-l)^{k-l} e^l l!} = C''' \frac{e^{-(k-l)\log(1-l/k)}}{e^l l!}.$$

As for every $x \in [0,1[$, $-\log(1-x) \leq x/(1-x)$, then $-(k-l)\log(1-l/k) \leq l$. As a consequence

$$\binom{k}{l} k^{-l} \leq C''' \frac{1}{l!}. \tag{7.17}$$

Then, there exists $D' > 0$ such that

$$\limsup_{n\to\infty} \frac{\mathbb{P}_k(N_n = l)}{k^\alpha \log(k) n^{-3/2}\gamma^n} \leq D'\big[\frac{1}{l^{1+\alpha}} + \frac{1}{l!} + \binom{k}{l} l^{-\alpha}(l/k)^l (1-l/k)^{k-l}\big].$$

Then,

$$\begin{aligned}
\limsup_{n\to\infty} \frac{\mathbb{P}_k(N_n = l)}{k^\alpha \log(k) n^{-3/2}\gamma^n} &= \limsup_{n\to\infty} \sum_{l'=l}^{k} \frac{\mathbb{P}_k(N_n = l')}{k^\alpha \log(k) n^{-3/2}\gamma^n} \\
&= \sum_{l'=l}^{k} \limsup_{n\to\infty} \frac{\mathbb{P}_k(N_n = l')}{k^\alpha \log(k) n^{-3/2}\gamma^n} \\
&\leq D \sum_{l'=l}^{k} \big[\frac{1}{l'^{1+\alpha}} + \frac{1}{l'!} + \binom{k}{l'} l'^{-\alpha}(l'/k)^{l'} (1-l'/k)^{k-l'}\big] \\
&\leq D\left[\sum_{l'=l}^{k} \big[\frac{1}{l'^{1+\alpha}} + \frac{1}{l'!}\big] + l'^{-\alpha}\right]
\end{aligned}$$

Recalling that $\mathbb{P}_k(Z_n > 0) \sim c\alpha_k n^{-3/2}\gamma^n$, $(n \to \infty)$ and $\alpha_k \geq M_{\text{-}} \log(k) k^\alpha$, $(k \in \mathbb{N})$ (see Theorem 7.2.2), we have

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P}_k(N_n \geq l \mid Z_n > 0) \\
= \; &\limsup_{n\to\infty} \frac{\mathbb{P}_k(N_n \geq l)}{c\alpha_k n^{-3/2}\gamma^n}
\end{aligned}$$

$$\leq \quad (cM_{\text{-}})^{-1} D \Big[\sum_{l'=l}^{k} \big[\frac{1}{l^{1+\alpha}} + \frac{1}{l!} \big] + l^{-\alpha} \Big].$$

This gives the first inequality of the proposition with

$$A_l = (cM_{\text{-}})^{-1} D \big[\sum_{l'=l}^{\infty} \big[\frac{1}{l^{1+\alpha}} + \frac{1}{l!} \big] + l^{-\alpha} \big].$$

We can prove similarly the lower bound. By Lemma 7.3.1, for every $x > 0$,

$$\mathbb{P}(p(\mathbf{f}_n) \geq x) \geq \mathbb{P}(L_n \geq \log(x\mu))/4.$$

Then, using also (7.3.1), for all $0 \leq l < k$ and $N > 0$,

$$\begin{aligned} \mathbb{P}(p(\mathbf{f}_n) \in [l/k, Nl/k[) &= \mathbb{P}(p(\mathbf{f}_n) \geq l/k) - \mathbb{P}(p(\mathbf{f}_n) \geq Nl/k) \\ &\geq \mathbb{P}(L_n \geq \log(\mu l/k))/4 - \mathbb{P}(\exp(L_n) \geq Nl/k). \end{aligned}$$

By (7.20) , we get

$$\liminf_{n\to\infty} \frac{\mathbb{P}(p(\mathbf{f}_n) \in [l/k, Nl/k[)}{n^{-3/2}\gamma^n} \geq (k/l)^{\alpha} [\mu^{-\alpha} u(\log(k) - \log(\mu l))/4 - N^{-\alpha} u(\log(k) - \log(Nl))].$$

Then, as u is linearly growing, we can fix $N \geq 1$ so that there exists $C > 0$ such that

$$\liminf_{k\to\infty} \liminf_{n\to\infty} \frac{\mathbb{P}(p(\mathbf{f}_n) \in [l/k, Nl/k[)}{k^{\alpha} \log(k) n^{-3/2}\gamma^n} \geq l^{-\alpha} C. \tag{7.18}$$

Using that

$$\mathbb{P}_k(N_n = l) = \int_0^1 \mathbb{P}(p(\mathbf{f}_n) \in \mathrm{d}x) \binom{k}{l} x^l (1-x)^{k-l},$$

and $x \to x^l(1-x)^{k-l}$ decreases on $[l/k, 1]$, we have, for every $k \geq Nl$,

$$\mathbb{P}_k(N_n = l) \geq \mathbb{P}(p(\mathbf{f}_n) \in [l/k, Nl/k[) \binom{k}{l} (Nl/k)^l (1 - Nl/k)^{k-l}.$$

Then (7.18) and $\lim_{k\to\infty} \binom{k}{l} (Nl/k)^l (1 - Nl/k)^{k-l} > 0$ ensures that

$$\liminf_{k\to\infty} \liminf_{n\to\infty} \frac{\mathbb{P}_k(N_n = l)}{k^{\alpha} \log(k) n^{-3/2}\gamma^n} > 0.$$

Use $\mathbb{P}_k(Z_n > 0) \sim c\alpha_k n^{-3/2}\gamma^n$ and the upperbound of α_k given in Theorem 7.2.2 to conclude. $\square$

7.3.3 Proofs of Section 7.2.3

Proof of Theorem 7.2.5. In the (WS+IS) case, recall that (see Section 3.1),

$$\mathbb{P}_k(Z_n^{(1)} > 0,\ Z_n^{(2)} > 0) = \mathbb{E}(p(\mathbf{f}_n)^2).$$

Thus, for every $\epsilon > 0$,

$$\mathbb{P}_k(Z_n^{(1)} > 0,\ Z_n^{(2)} > 0 \mid Z_n > 0) \geq \epsilon^2 \mathbb{P}_k(p(\mathbf{f}_n) \geq \epsilon \mid Z_n > 0).$$

By Proposition 7.2.3, we get

$$\mathbb{P}_k(p(\mathbf{f}_n) \geq \epsilon \mid Z_n > 0) \stackrel{n\to\infty}{\longrightarrow} 0.$$

In the (WS) case, by (7.11), for every $\epsilon \in]0,1]$:

$$\mathbb{P}_k(Z_n > 0) \quad \geq \quad \int_0^\epsilon \mathbb{P}(p(\mathbf{f}_n) \in \mathrm{d}x)(1-(1-x)^k).$$

Moreover

$$\begin{aligned}
& \Big| \int_0^\epsilon \mathbb{P}(p(\mathbf{f}_n) \in \mathrm{d}x)(1-(1-x)^k) - \int_0^\epsilon \mathbb{P}(p(\mathbf{f}_n) \in \mathrm{d}x)kx \Big| \\
\leq \quad & k \sup_{x\in[0,\epsilon[} \Big\{ \Big| \frac{1-(1-x)^k}{kx} - 1 \Big| \Big\} \int_0^\epsilon \mathbb{P}(p(\mathbf{f}_n) \in \mathrm{d}x)x \\
\leq \quad & k \sup_{x\in[0,\epsilon[} \Big\{ \Big| \frac{1-(1-x)^k}{kx} - 1 \Big| \Big\} \mathbb{P}_1(Z_n > 0).
\end{aligned}$$

Putting these two inequalities together yields

$$\mathbb{P}_k(Z_n > 0) \geq k \int_0^\epsilon \mathbb{P}(p(\mathbf{f}_n) \in \mathrm{d}x)x - k \sup_{x\in[0,\epsilon[} \Big\{ \Big| \frac{1-(1-x)^k}{kx} - 1 \Big| \Big\} \mathbb{P}_1(Z_n > 0).$$

Then

$$\begin{aligned}
\mathbb{P}_1(p(\mathbf{f}_n) \in [0,\epsilon[),\ Z_n > 0) \quad &= \quad \int_0^\epsilon \mathbb{P}(p(\mathbf{f}_n) \in \mathrm{d}x)x \\
&\leq \quad \mathbb{P}_k(Z_n > 0)/k + \sup_{x\in[0,\epsilon[} \Big\{ \Big| \frac{1-(1-x)^k}{kx} - 1 \Big| \Big\} \mathbb{P}_1(Z_n > 0).
\end{aligned}$$

Dividing by $\mathbb{P}_1(Z_n > 0)$ and letting $n \to \infty$ ensure that

$$\begin{aligned}
& \limsup_{n\to\infty} \mathbb{P}_1(p(\mathbf{f}_n) \in [0,\epsilon[) \mid Z_n > 0) \\
\leq \quad & \limsup_{n\to\infty} \frac{\mathbb{P}_k(Z_n > 0)}{k\mathbb{P}_1(Z_n > 0)} + \sup_{x\in[0,\epsilon[} \Big\{ \Big| \frac{1-(1-x)^k}{kx} - 1 \Big| \Big\}
\end{aligned}$$

$$\leq \quad \frac{\alpha_k}{k} + \sup_{x\in[0,\epsilon[} \left\{ \left| \frac{1-(1-x)^k}{kx} - 1 \right| \right\}.$$

Finally recall Theorem 7.2.2 and use

$$\alpha_k/k \overset{k\to\infty}{\longrightarrow} 0, \qquad \forall k \in \mathbb{N}^*, \quad \sup_{x\in[0,\epsilon[} \left\{ \left| \frac{1-(1-x)^k}{kx} - 1 \right| \right\} \overset{\epsilon\to 0}{\longrightarrow} 0,$$

to get $\lim_{\epsilon\to 0+} \limsup_{n\to\infty} \mathbb{P}_k(p(\mathbf{f}_n) \leq \epsilon \mid Z_n > 0) = 0$. □

Proof of Proposition 7.2.6. Recall that for every $\mathbf{g}_n \in \mathcal{F}^n$, $\mathbb{P}_k(Z_{\mathbf{g}_n} > 0) = 1-(1-p(\mathbf{g}_n))^k$. Thus,

$$\begin{aligned}
\mathbb{P}_k(p(\mathbf{f}_n) \in \mathrm{d}x \mid Z_n > 0) &= \frac{\mathbb{P}(p(\mathbf{f}_n) \in \mathrm{d}x)(1-(1-x)^k)}{\mathbb{P}_k(Z_n > 0)} \\
&= \mathbb{P}_1(p(\mathbf{f}_n) \in \mathrm{d}x \mid Z_n > 0)\frac{\mathbb{P}_1(Z_n > 0)}{\mathbb{P}_k(Z_n > 0)}\frac{(1-(1-x)^k)}{x}.
\end{aligned}$$

Then, for every $\epsilon > 0$,

$$\begin{aligned}
& \limsup_{n\to\infty} \mathbb{P}_k(p(\mathbf{f}_n) \geq \epsilon \mid Z_n > 0) \\
= \quad & \frac{1}{\alpha_k} \limsup_{n\to\infty} \int_\epsilon^1 \mathbb{P}_1(p(\mathbf{f}_n) \in \mathrm{d}x \mid Z_n > 0)\frac{(1-(1-x)^k)}{x} \\
\leq \quad & \frac{1}{\epsilon\alpha_k},
\end{aligned}$$

and the left hand part tends to zero as k tends to infinity by Theorem 7.2.2. This ends up the proof. □

7.3.4 Proofs of section 7.2.4

To prove Theorem 7.2.7, we first prove that the probability generating function G_k of the quasistationary distributions Υ_k verify the same functional equation. And we prove that in the (SS+IS) case, the quasistationary distributions do not depend on k. Then we prove a lemma which ensures the uniqueness of the solution of this functional equation in the (SS) case.

Lemma 7.3.2. *In the subcritical case, the generating function G_k of Υ_k verifies*

$$\mathbb{E}(G_k(f(s))) = \gamma G_k(s) + 1 - \gamma, \qquad G_k(0) = 0.$$

In the (IS+SS) case, for every $k \geq 1$, $\Upsilon_k = \Upsilon_1$. In the (SS) case, $G_1'(1) < \infty$.

Proof. Let f_0 be distributed as f and independent of $(Z_n)_{n\in\mathbb{N}}$. For every $n \in \mathbb{N}$,

$$\begin{aligned}
1-\mathbb{E}_k(s^{Z_{n+1}} \mid Z_{n+1}>0) &= \frac{\mathbb{E}_k(1-s^{Z_{n+1}})}{\mathbb{P}_k(Z_{n+1}>0)} \\
&= \frac{1}{\mathbb{P}_k(Z_{n+1}>0)}\sum_{i=1}^{\infty}\mathbb{P}_k(Z_n=i)\mathbb{E}_k(1-s^{Z_{n+1}} \mid Z_n=i) \\
&= \frac{\mathbb{P}_k(Z_n>0)}{\mathbb{P}_k(Z_{n+1}>0)}\sum_{i=1}^{\infty}\mathbb{P}_k(Z_n=i \mid Z_n>0)\mathbb{E}(1-f_0(s)^i) \\
&= \frac{\mathbb{P}_k(Z_n>0)}{\mathbb{P}_k(Z_{n+1}>0)}\mathbb{E}_k(1-f_0^{Z_n}(s) \mid Z_n>0).
\end{aligned}$$

Then letting n tend to infinity and using asymptotics given in Preliminaries gives

$$1-G_k(s)=\gamma^{-1}\mathbb{E}(1-G_k(f_0(s))),$$

where $\gamma=\mathbb{E}(f'(1))$ in the (SS+IS) case. This gives the equation of the lemma.

In the (SS) case, the fact that $G_k'(1)<\infty$ is proved in [43] for $k=1$. The proof can be generalized to $k\geq 1$. And we can then use the uniqueness of the solution of the functional equation given below to prove that for every $k\geq 1$, $G_k=G_1$.

But in the (SS+IS) case, we can also directly prove uniqueness of all quasistationary distributions. Indeed, for every $i\geq 1$, $\mathbb{P}_2(Z_n=i)$ is equal to

$$\mathbb{P}(Z_n^{(1)}=i,\ Z_n^{(2)}=0)+\mathbb{P}(Z_n^{(1)}=0,\ Z_n^{(2)}=i)+\mathbb{P}_2(Z_n=i,\ Z_n^{(1)}>0,\ Z_n^{(2)}>0).$$

Moreover $|\mathbb{P}(Z_n^{(1)}=i,\ Z_n^{(2)}=0)-P_1(Z_n=i)|\leq \mathbb{P}(Z_n^{(1)}>0,\ Z_n^{(2)}>0)$, then

$$|\mathbb{P}_2(Z_n=i)-2\mathbb{P}_1(Z_n=i)|\leq 3\mathbb{P}(Z_n^{(1)}>0,\ Z_n^{(2)}>0).$$

Thus, using Proposition 7.2.3,

$$\lim_{n\to\infty}\frac{\mathbb{P}_2(Z_n=i)}{\mathbb{P}_2(Z_n>0)}=\lim_{n\to\infty}\frac{2\mathbb{P}_1(Z_n=i)}{\mathbb{P}_2(Z_n>0)}.$$

As $\alpha_2=\lim_{n\to\infty}\mathbb{P}_2(Z_n>0)/\mathbb{P}_1(Z_n>0)=2$, we have

$$\begin{aligned}
\mathbb{P}(\Upsilon_2=i) &= \lim_{n\to\infty}\mathbb{P}_2(Z_n=i \mid Z_n>0) \\
&= \lim_{n\to\infty}\frac{2\mathbb{P}_1(Z_n=i \mid Z_n>0)\mathbb{P}_1(Z_n>0)}{\mathbb{P}_2(Z_n>0)} \\
&= \mathbb{P}(\Upsilon_1=i).
\end{aligned}$$

Then $\Upsilon_1 \overset{d}{=} \Upsilon_2$ and the same argument ensures that for every $k\geq 1$, $\Upsilon_k=\Upsilon_1$. $\square$

To prove the uniqueness of the functional equation in the (SS) case, we need the following result.

Lemma 7.3.3. *If* $H : [0,1] \to \mathbb{R}$ *is a power series continuous on* $[0,1]$, $H(1) = 0$ *and*

$$H(s) = \frac{\mathbb{E}(H(f(s))f'(s))}{\mathbb{E}(f'(1))}, \quad (0 \le s \le 1), \tag{7.19}$$

then $H \equiv 0$.

Proof. FIRST CASE : There exists $s_0 \in [0,1[$ such that $\mathbb{E}(f'(s_0)) = \mathbb{E}(f'(1))$.

The monotonicity of f' implies

$$f'(s_0) = f'(1) \quad \text{a.s.},$$

and f' is a.s. constant on $[s_0, 1]$. As it is a power series, f' is a.s. constant. Thus

$$f(s) = f'(1)s + (1 - f'(1)) \quad (0 \le s \le 1), \qquad f'(1) \le 1 \quad \text{a.s.}$$

Moreover, let $|H(\alpha)| = \sup\{|H(s)|,\ s \in [0,1]\}$ with $\alpha \in [0,1[$, and note that

$$\mathbb{E}\big(f'(1)(H(\alpha) - H(f(\alpha)))\big) = 0.$$

Thus $H(f(\alpha)) = H(\alpha)$ a.s. and by induction, recalling that $F_n = f_0 \circ f_1 \cdots \circ f_{n-1}$, we have

$$H(F_n(\alpha)) = H(\alpha) \quad \text{a.s.}$$

The orbit of $(F_n(\alpha))_{n\in\mathbb{N}}$ has a point of accumulation at 1, since $\alpha < 1$ and Z_n is subcritical. As H is a power series, then H is constant and equals to zero since $H(1) = 0$.

SECOND CASE : For every $s_0 \in [0,1[,\ \mathbb{E}(f'(s_0)) < \mathbb{E}(f'(1))$.
If $H \neq 0$ then there exists $\alpha \in [0,1[$ such that

$$\sup\{\mid H(s)) \mid : s \in [0,\alpha]\} > 0$$

Let $\alpha_n \in [\alpha, 1[$ such that $\alpha_n \overset{n\to\infty}{\longrightarrow} 1$. Then, for every $n \in \mathbb{N}$, there exists $\beta_n \in [0, \alpha_n]$ such that :

$$\begin{aligned}
\sup\{\mid H(s) \mid : s \in [0,\alpha_n]\} &= \mid H(\beta_n) \mid \\
&\le \frac{\mathbb{E}(f'(\beta_n))}{\mathbb{E}(f'(1))} \sup\{|H(s)|,\ 0 \le s \le 1\} \\
&< \sup\{\mid H(s) \mid,\ 0 \le s \le 1\},
\end{aligned}$$

since $\sup\{\mid H(s) \mid,\ 0 \le s \le 1\} > 0$ and $\mathbb{E}(f'(\beta_n)) < \mathbb{E}(f'(1))$. As $I \cap J = \varnothing$, $\sup I < \sup(I \cup J) \Rightarrow \sup I < \sup J$, we get

$$\sup\{\mid H(s) \mid : s \in [0,\alpha_n]\} < \sup\{\mid H(s) \mid : s \in]\alpha_n, 1]\}.$$

And $H(s) \overset{s\to 1}{\longrightarrow} 0$ leads to a contradiction letting $n \to \infty$. So $H = 0$. $\square$

We can now easily prove the uniqueness in the (SS) case in Theorem 7.2.7.

Lemma 7.3.4. *There exists at most one probability generating function G satisfying*

$$\mathbb{E}(G(f(s))) = \mathbb{E}(f'(1))G(s) + 1 - \mathbb{E}(f'(1)) \quad (0 \leq s \leq 1), \quad G(0) = 0, \quad G'(1) < \infty.$$

Proof. Assume that G_1 and G_2 are two probability generating functions which verify the equation above. By differentiation, G_1' and G_2' satisfy

$$\mathbb{E}(G'(f(s))f'(s)) = \mathbb{E}(f'(1))G'(s).$$

Then $H := G_2'(1)G_1' - G_1'(1)G_2'$ verifies the conditions of Lemma 7.3.3. As a consequence,

$$G_2'(1)G_1' = G_1'(1)G_2'.$$

And $G_1(0) = G_2(0) = 0$, $G_2(1) = G_1(1) = 1$ ensure that $G_1 = G_2$, which give the uniqueness. □

Finally, we prove that if $Z_1 \in \{0, 1, N\}$ for some $N \geq 1$, then $\Upsilon_1 \stackrel{d}{=} \Upsilon_N$.

Proof. For every $s \in [0, 1]$, we have

$$\begin{aligned}
\mathbb{E}_1(s^{Z_{n+1}} \mid Z_{n+1} > 0) &= \frac{\mathbb{E}_1\big(\mathbb{1}(Z_1 = 1,\ Z_{n+1} > 0)s^{Z_{n+1}} + \mathbb{1}(Z_1 = N,\ Z_{n+1} > 0)s^{Z_{n+1}}\big)}{\mathbb{P}_1(Z_{n+1} > 0)} \\
&= \frac{\mathbb{P}_1(Z_1 = 1)\mathbb{P}_1(Z_n > 0)}{\mathbb{P}_1(Z_{n+1} > 0)}\mathbb{E}_1(s^{Z_n} \mid Z_n > 0) \\
&\quad + \frac{\mathbb{P}_1(Z_1 = N)\mathbb{P}_N(Z_n > 0)}{\mathbb{P}_1(Z_{n+1} > 0)}\mathbb{E}_N(s^{Z_n} \mid Z_n > 0).
\end{aligned}$$

For every $k \in \mathbb{N}$, letting $n \to \infty$ using (7.1) yields

$$\mathbb{E}(s^{\Upsilon_1}) = \frac{\mathbb{P}_1(Z_1 = 1)}{\gamma}\mathbb{E}(s^{\Upsilon_1}) + \frac{\mathbb{P}_1(Z_1 = N)\alpha_N}{\gamma}\mathbb{E}(s^{\Upsilon_N}),$$

which proves $\Upsilon_1 \stackrel{d}{=} \Upsilon_N$. □

7.3.5 Proof of Section 7.2.5

Proof of Proposition 7.2.8. First, we have

$$\mathbb{P}_k(Z_1 = l_1, ..., Z_n = l_n | Z_{n+p} > 0) = \mathbb{P}_k(Z_1 = l_1, ..., Z_n = l_n) \frac{\mathbb{P}_{l_n}(Z_p > 0)}{\mathbb{P}_k(Z_{n+p} > 0)}.$$

Then, using (6.2), (7.3), (7.4), we get

$$\lim_{p\to\infty} \mathbb{P}_k(Z_1 = l_1, ..., Z_n = l_n | Z_{n+p} > 0) = \gamma^{-n} \frac{\alpha_{l_n}}{\alpha_k} \mathbb{P}_k(Z_1 = l_1, ..., Z_n = l_n).$$

and recall $\alpha_l = l$ in the (SS+IS) case to get the distribution of $(Y_n)_{n\in\mathbb{N}}$.

To get the limit distribution of $(Y_n)_{n\in\mathbb{N}}$, note that, for every $l \in \mathbb{N}^*$,

$$\mathbb{P}_k(Y_n = l) = \gamma^{-n} \frac{\alpha_l}{\alpha_k} \mathbb{P}_k(Z_n = l) = \gamma^{-n} \mathbb{P}_k(Z_n > 0) \frac{\alpha_l}{\alpha_k} \mathbb{P}_k(Z_n = l \mid Z_n > 0).$$

Use respectively (6.2) and (7.3) to get the limit in distribution in the (SS) case and the (IS).

Finally, in the (WS) case, by (7.4), there exists $C > 0$ such that

$$\begin{aligned} \mathbb{P}_k(Y_n \leq l) &\leq Cn^{-3/2} \frac{\alpha_l}{\alpha_k} \mathbb{P}_k(Z_n \leq l \mid Z_n > 0) \\ &\leq Cn^{-3/2} \frac{\alpha_l}{\alpha_k}. \end{aligned}$$

Then Borel-Cantelli Lemma ensures that Y_n tends a.s. to infinity as $n \to \infty$. □

Proof of (7.7). To prove the convergence and the equality, note that

$$\begin{aligned} \mathbb{P}_k\big(\mathbf{f}_p \in \mathrm{d}\mathbf{g}_p | Z_{n+p} > 0\big) &= \frac{\mathbb{P}\big(\mathbf{f}_p \in \mathrm{d}\mathbf{g}_p\big) \mathbb{E}_k(\mathbb{P}_{Z_{\mathbf{g}_p}}(Z_n > 0))}{\mathbb{P}_k(Z_{n+p} > 0)} \\ &= \frac{\mathbb{P}_1(Z_n > 0)}{\mathbb{P}_k(Z_{n+p} > 0)} \sum_{l=1}^{\infty} \mathbb{P}_k(Z_{\mathbf{g}_p} = l) \frac{\mathbb{P}_l(Z_n > 0)}{\mathbb{P}_1(Z_n > 0)}. \end{aligned}$$

Asymptotics given in Introduction ensure that

$$\frac{\mathbb{P}_1(Z_n > 0)}{\mathbb{P}_k(Z_{n+p} > 0)} \overset{n\to\infty}{\longrightarrow} \frac{1}{\gamma^p \alpha_k},$$

and using the bounded convergence Theorem with

$$\frac{\mathbb{P}_l(Z_n > 0)}{\mathbb{P}_1(Z_n > 0)} \overset{n\to\infty}{\longrightarrow} \alpha_l, \quad \frac{\mathbb{P}_l(Z_n > 0)}{\mathbb{P}_1(Z_n > 0)} \leq l, \quad \mathbb{E}(Z_{\mathbf{g}_p}) < \infty.$$

ensures that

$$\lim_{n\to\infty} \mathbb{P}_k\big(\mathbf{f}_p \in \mathrm{d}\mathbf{g}_p | Z_{n+p} > 0\big) = \gamma^{-p}\mathbb{P}\big(\mathbf{f}_p \in \mathrm{d}\mathbf{g}_p\big) \sum_{l=1}^{\infty} \mathbb{P}_k(Z_{\mathbf{g}_p} = l)\frac{\alpha_l}{\alpha_k}.$$

This completes the proof. □

7.4 Appendix : Random walk with negative drift

We study here the random walk $(S_n)_{n\in\mathbb{N}}$ with negative drift. Indeed, in the linear fractional case, the survival probability is a functional of the random walk obtained by summing the successive means of environments (see (6.6). In the general case, the random walk appears in the lowerbound of the survival probability (see (7.10)). More precisely, we need to control the successive values of the random walk with negative drift conditioned to stay above $-x < 0$.

More specifically, let $(X_i)_{i\in\mathbb{N}}$ iid random variables distributed as X with

$$\mathbb{E}(X) < 0.$$

We assume that for every $z \in [0,1]$, $\mathbb{E}(\exp(zX)) < \infty$ and $\mathbb{E}(X\exp(\alpha X)) = 0$ for some $0 < \alpha < 1$. Set $\gamma := \mathbb{E}(\exp(\alpha X))$,

$$S_n := \sum_{i=0}^{n-1} X_i, \qquad (S_0 = 0),$$

and for all $n \in \mathbb{N},\ k \in \mathbb{N}$,

$$L_n = \min\{S_i,\ 0 \le i \le n\}.$$

Its asymptotic is given in Lemma 4.1 in [43] or Lemma 7 in [50]. There exists a linearly increasing positive function u such that, as $n \to \infty$

$$\mathbb{P}(L_n \ge -x) \sim e^{\alpha x} u(x) n^{-3/2}\gamma^n, \tag{7.20}$$

for $x \ge 0$ if the distribution X is non-lattice, and for $x \in \lambda\mathbb{Z}$ if the distribution of X is supported by a centered lattice $\lambda\mathbb{Z}$.
Moreover for each $\theta > \alpha$, there exists $c_\theta > 0$ such that

$$\mathbb{P}(L_n \ge -x) \le c_\theta e^{\theta x} n^{-3/2}\gamma^n, \qquad (x \ge 0,\ n \in \mathbb{N}). \tag{7.21}$$

Finally, using (7.20) and the fact that u grows linearly, there exist $c_-, c_+ > 0$ such that the two following positive measures on $[0,1]$,

$$\nu_-(\mathrm{d}x) = c_- \log(1/x)x^{-\alpha-1}\mathrm{d}x, \qquad \nu_+(\mathrm{d}x) = c_+(\delta_1(\mathrm{d}x) + \log(1/x)x^{-\alpha-1}\mathrm{d}x),$$

verify for every $x \in]0,1]$

$$\nu_-([x,1]) \leq \lim_{n\to\infty} \frac{\mathbb{P}(e^{L_n} \geq x)}{n^{-3/2}\gamma^n} \leq \nu_+([x,1]). \tag{7.22}$$

We need to control the successive values of the random walk conditioned to stay above $-x$ ($x \geq 0$). Under integrability conditions, it is known that the process $(S_{[nt]}/n^{1/2}|L_n \geq 0)$ converges weakly to Brownian meander as $n \to \infty$ (see [53]). Moreover Durrett [34] has proved that if there exists $q > 2$ such that $P\{X_1 > x\} \sim x^{-q}L(x)$ as $x \to \infty$, where L is slowly varying, then $(S_{[nt]}/n|L_n \geq 0)$ converges weakly to a non degenerate limit with a single jump.
We prove here that the random walk conditioned to stay above $-x$ ($x \geq 0$) spends very few time close to its minimum, by giving an upperbound of the number of visits of a level of the random walk reflected on its minimum. To be more specific, define

$$N_n(k) = \mathrm{card}\{i \in \mathbb{N},\ i \leq n,\ k \leq S_i - L_n < k+1\}.$$

Lemma 7.4.1. *For every $\theta > \alpha$, there exists $d > 0$ such that*

$$\limsup_{n\to\infty} \mathbb{P}(N_n(k) \geq l \mid L_n \geq -x) \leq de^{\theta k}/\sqrt{l}, \qquad (k,l \in \mathbb{N},\ x \geq 0).$$

Moreover for all $\theta > \alpha$ and $x \geq 0$, there exists $C > 0$ such that

$$\mathbb{P}(N_n(k) \geq l \mid L_n \geq -x) \leq Ce^{\theta k}/\sqrt{l}, \quad (k,n,l \in \mathbb{N}). \tag{7.23}$$

Moreover, we will use the following consequence of the preceding lemma.

Corollary 7.4.2. *If $\alpha < 1/2$, there exists $\beta > 0$ such that for all $x \geq 0$ and $n \in \mathbb{N}$,*

$$\mathbb{P}\big(\sum_{i=0}^{n} \exp(L_n - S_i) \leq \beta \mid L_n \geq -x\big) \geq 1/4.$$

For the sake of simplicity, we assume that $X \in \mathcal{Z}$ a.s. for the proof of Lemma 7.4.1. Thus

$$\forall k,n \in \mathbb{N}^2, \quad N_n(k) = \mathrm{card}\{i \in \mathbb{N},\ i \leq n,\ S_i - L_n = k\},$$

and we denote by $(T_j\ :\ 1 \leq j \leq N_n(k))$ the successive times before n when $(S_i - L_n)_{i\in\mathbb{N}}$ visits k. That is

$$T_1 = \inf\{0 \leq i \leq n\ :\ S_i - L_n = k\}, \quad T_{j+1} = \inf\{T_j < i \leq n\ :\ S_i - L_n = k\}.$$

First, cutting the path of the random walk between two of these passage times enables us to prove the following result.

Lemma 7.4.3. *If $X \in \mathcal{Z}$ a.s., then for all n, k, l, i and $0 \le h \le n$, we have*

$$\mathbb{P}(L_n \ge -i,\ N_n(k) \ge 2l,\ T_l + n - T_{N_n(k)} = h) \le (k+1)\mathbb{P}(L_{n-h} \ge -k)\mathbb{P}(L_h \ge -i),$$

and

$$\mathbb{P}(L_n \ge -i,\ N_n(k) \ge 2l,\ T_1 + n - T_l = h) \le (k+1)\mathbb{P}(L_{n-h} \ge -k)\mathbb{P}(L_h \ge -i).$$

Proof. We introduce M_n the first reaching time of the minimum L_n before time n and $R_n(l)$ the last passage time of l before time n

$$M_n = \inf\{j \in [1,n] : S_j = L_n\}, \qquad R_n(l) := \sup\{j \in [1,n] : S_j = l\}.$$

First, we consider the case where $M_n \in [0, T_l] \cup [T_{N_n(k)}, n]$ and split the path of the random walk between times T_l and $T_{N_n(k)}$. For all $j \le 0$, $k \ge 0$ and $0 \le n_1 < n_2 \le n$, introduce then

$$\begin{aligned}
A(j,n_1,n_2) &= \{L_n = j,\ N_n(k) \ge 2l,\ T_l = n_1,\ T_{N_n(k)} = n_2,\ M_n \in [0,n_1] \cup [n_2,n]\},\\
B(j,n_1,n_2) &= \{\forall m \in [1,n_1] :\ S_m \ge j,\ S_{n_1} = S_{n_2} = j+k,\\
&\qquad \forall m \in [n_2+1,n] : S_m \ge j,\ S_m \ne j+k,\\
&\qquad \exists a \in [0,n_1] \cup [n_2,n],\ S_a = j\},\\
C(j,n_1,n_2) &= \{\forall m \in [n_1,n_2] : S_m \ge j,\ S_{n_1} = S_{n_2} = j+k\}.
\end{aligned}$$

Note that conditionally on $D(n_1,n_2) := \{S_{n_1} = S_{n_2} = j+k\}$, $B(j,n_1,n_2)$ and $C(j,n_1,n_2)$ are independent,

$$\mathbb{P}(C(j,n_1,n_2) \mid S_{n_1} = j+k) \le \mathbb{P}(L_{n_2-n_1} \ge -k),$$

and

$$A(j,n_1,n_2) \subset B(j,n_1,n_2) \cap C(j,n_1,n_2).$$

Then, noting also that

$$\mathbb{P}(C(j,n_1,n_2) \mid D(n_1,n_2)) = \mathbb{P}(C(j,n_1,n_2) \mid S_{n_1} = j+k)\mathbb{P}(S_{n_1} = j+k)/\mathbb{P}(D(n_1,n_2)),$$

we have

$$\begin{aligned}
&\mathbb{P}(A(j,n_1,n_2))\\
\le\ & \mathbb{P}(D(n_1,n_2))\mathbb{P}(B(j,n_1,n_2) \mid D(n_1,n_2))\mathbb{P}(C(j,n_1,n_2) \mid D(n_1,n_2))\\
=\ & \mathbb{P}(S_{n_1} = j+k)\mathbb{P}(B(j,n_1,n_2) \mid D(n_1,n_2))\mathbb{P}(C(j,n_1,n_2) \mid S_{n_1} = j+k)\\
\le\ & \mathbb{P}(L_{n_2-n_1} \ge -k)\mathbb{P}(S_{n_1} = j+k)\mathbb{P}(B(j,n_1,n_2) \mid D(n_1,n_2)). \qquad (7.24)
\end{aligned}$$

Moreover,

$$\begin{aligned}
\{L_n \ge -i,\ N_n(k) \ge 2l,\ T_l + n - T_{N_n(k)} = h,\ M_n \in [0,T_l] \cup [T_{N_k(n)}, n]\}&\\
= \bigsqcup_{\substack{j \ge -i,\\ 1 \le n_1 < n_2 \le n,\ n_1+n-n_2 = h}} A(j,n_1,n_2).&
\end{aligned}$$

Then, using the last two relations,

$$\begin{aligned}
& \mathbb{P}(L_n \geq -i,\ N_n(k) \geq 2l,\ T_l + n - T_{N_n(k)} = h,\ M_n \in [0, T_l] \cup [T_{N_k(n)}, n]) \\
= & \sum_{\substack{j \geq -i, \\ 1 \leq n_1 < n_2 \leq n,\ n_2 - n_1 = n-h}} \mathbb{P}(A(j, n_1, n_2)) \\
\leq & \ \mathbb{P}(L_{n-h} \geq -k) \sum_{\substack{j \geq -i, \\ 1 \leq n_1 < n_2 \leq n,\ n_1 + n - n_2 = h}} \mathbb{P}(S_{n_1} = j + k)\mathbb{P}(B(j, n_1, n_2) \mid D(n_1, n_2)).
\end{aligned}$$

Concatenating the path of the random walk before time n_1 and after time n_2 gives

$$\begin{aligned}
& \mathbb{P}(L_n \geq -i,\ N_n(k) \geq 2l,\ T_l + n - T_{N_n(k)} = h,\ M_n \in [0, T_l] \cup [T_{N_k(n)}, n]) \\
\leq & \ \mathbb{P}(L_{n-h} \geq -k) \sum_{\substack{j \geq -i, \\ 1 \leq n_1 < n_2 \leq n,\ n_1 + n - n_2 = h}} \mathbb{P}(L_{n_1 + n - n_2} = j,\ R_{n_1 + n - n_2}(j + k) = n_1) \\
\leq & \ \mathbb{P}(L_{n-h} \geq -k) \sum_{j \geq -i} \mathbb{P}(L_h = j) \\
= & \ \mathbb{P}(L_{n-h} \geq -k)\mathbb{P}(L_h \geq -i). \qquad (7.25)
\end{aligned}$$

Second, we consider the case where $M_n \in [T_l, T_{N_n(k)}]$ and split the path of the random walk between times T_1 and T_l; For all $j, j' \leq 0$, $k \geq 0$ and $0 \leq n_1 < n_2 \leq n$, introduce then

$$\begin{aligned}
A'(j, n_1, n_2) &= \{L_n = -j,\ N_n(k) \geq 2l,\ T_l = n_1,\ T_{N_n(k)} = n_2,\ M_n \in [n_1, n_2]\}, \\
B'(j, j', n_1, n_2) &= \{\forall m \in [1, n_1] :\ S_m \geq j',\ S_{n_1} = S_{n_2} = j + k, \\
& \qquad \forall m \in]n_2, n] : S_m \geq j',\ S_m \neq j + k, \\
& \qquad \exists a \in [0, n_1] \cup [n_2, n] :\ S_a = j'\}, \\
C'(j, n_1, n_2) &= \{\forall m \in [n_1, n_2] : S_m \geq j,\ S_{n_1} = S_{n_2} = k + j, \\
& \qquad \exists a \in [n_1, n_2] : S_a = j\}.
\end{aligned}$$

Note that conditionally on $D(n_1, n_2) = \{S_{n_1} = S_{n_2} = j + k\}$, $B'(j, j', n_1, n_2)$ and $C'(j, n_1, n_2)$ are independent,

$$A'(j, n_1, n_2) \subset \bigsqcup_{j'=j}^{j+k} B'(j, j', n_1, n_2) \cap C'(j, n_1, n_2).$$

and we get the analogue of (7.24),

$$\mathbb{P}(A'(j, n_1, n_2)) \leq \sum_{j'=j}^{j+k} \mathbb{P}(L_{n_2 - n_1} \geq -k)\mathbb{P}(S_{n_1} = j + k)\mathbb{P}(B'(j, j', n_1, n_2) \mid D(n_1, n_2)).$$

Moreover

$$\{L_n \geq -i,\ N_n(k) \geq 2l,\ T_l + n - T_{N_n(k)} = h,\ M \in [T_l, T_{N_k(n)}]\}$$

$$= \bigsqcup_{\substack{j\geq -i,\\ 1\leq n_1<n_2\leq n,\ n_1+n-n_2=h}} A'(j,n_1,n_2).$$

Then, following the proof of (9.11), we get

$$\begin{aligned}
&\mathbb{P}(L_n \geq -i,\ N_n(k)\geq 2l,\ T_l+n-T_{N_n(k)}=h,\ M_n\in[T_l,T_{N_k(n)}])\\
\leq\ & \mathbb{P}(L_{n-h}\geq -k)\sum_{\substack{j'\geq -i,\\ j\in[j'-k,j']}}\ \sum_{\substack{1\leq n_1<n_2\leq n,\\ n_1+n-n_2=h}} \mathbb{P}(S_{n_1}=j+k)\mathbb{P}(B'(j,j',n_1,n_2)\mid D(n_1,n_2))\\
\leq\ & \mathbb{P}(L_{n-h}\geq -k)\sum_{j'\geq -i} k \max_{j\in[j'-k,j']} \sum_{\substack{1\leq n_1<n_2\leq n,\\ n_1+n-n_2=h}} \mathbb{P}(S_{n_1}=j+k)\times \qquad (7.26)\\
& \qquad\qquad \mathbb{P}(B'(j,j',n_1,n_2)\mid D(n_1,n_2))\\
\leq\ & \mathbb{P}(L_{n-h}\geq -k)\sum_{j'\geq -i} k\mathbb{P}(L_h=j')\\
\leq\ & k\mathbb{P}(L_{n-h}\geq -k)\mathbb{P}(L_h\geq -i). \qquad (7.27)
\end{aligned}$$

Combining the inequalities (9.11) and (9.12), we get

$$\mathbb{P}(L_n \geq -i,\ N_n(k)\geq 2l,\ T_l+n-T_{N_n(k)}=h) \leq (k+1)\mathbb{P}(L_{n-h}\geq -k)\mathbb{P}(L_h\geq -i),$$

which proves the first inequality of the lemma. The second one can be proved similarly concatenating the random walk between $[0,T_1]$ and $[T_{N_n(k)},n]$. □

Proof of Lemma 7.4.1. Let $h\in\mathbb{N}$ such that $h\geq n/2$. The first inequality of Lemma 7.4.3 above ensures that

$$\mathbb{P}(L_n \geq -i,\ N_n(k)\geq 2l,\ T_l+n-T_{N_n(k)}=h) \leq (k+1)\mathbb{P}(L_h\geq -i)\mathbb{P}(L_{n-h}\geq -k).$$

Using (7.21),

$$\mathbb{P}(L_n \geq -i,\ N_n(k)\geq 2l,\ T_l+n-T_{N_n(k)}=h) \leq c_\theta(k+1)\mathbb{P}(L_h\geq -i)e^{\theta k}(n-h)^{-3/2}\gamma^{n-h}.$$

Moreover, using (7.20), for every $i\in\mathbb{N}$, there exists $n_0\in\mathbb{N}$ such that for all $n_0/2\leq n/2\leq h$,

$$\mathbb{P}(L_h\geq -i)\leq 2e^{i\alpha}u(i)h^{-3/2}\gamma^{-h}\leq 2.2^{3/2}e^{i\alpha}u(i)n^{-3/2}\gamma^{h}. \qquad (7.28)$$

Then, writing $c'_\theta = 2.2^{3/2}.c_\theta$,

$$\mathbb{P}(L_n \geq -i,\ N_n(k)\geq 2l,\ T_l+n-T_{N_n(k)}=h) \leq c'_\theta e^{\alpha i}u(i)(k+1)e^{\theta k}\gamma^n n^{-3/2}(n-h)^{-3/2}. \qquad (7.29)$$

Similarly, for every h such that $n_0/2 \leq n/2 \leq h$, the second inequality of Lemma 7.4.3 above ensures that

$$\mathbb{P}(L_n \geq -i,\ N_n(k) \geq 2l,\ T_1+n-T_l = h) \leq c'_\theta e^{\alpha i} u(i)(k+1)e^{\theta k}\gamma^n n^{-3/2}(n-h)^{-3/2}. \tag{7.30}$$

Noting that a.s. $\{N_n(k) \geq 2l\}$ is equal to

$$\bigcup_{h=n/2}^{n-l} \{N_n(k) \geq 2l,\ T_l + n - T_{N_n(k)} = h\} \bigcup_{h=n/2}^{n-l} \{N_n(k) \geq 2l,\ T_1 + n - T_l = h\},$$

we can combine the last two inequalities (7.29) and (7.30), which give for every $n \geq n_0$,

$$\begin{aligned}
\mathbb{P}(L_n \geq -i,\ N_n(k) \geq 2l) &\leq \sum_{n/2 \leq h \leq n-l} \mathbb{P}(L_n \geq -i,\ N_n(k) \geq 2l,\ T_l + n - T_{N_n(k)} = h) \\
&\quad + \sum_{n/2 \leq h \leq n-l} \mathbb{P}(L_n \geq -i,\ N_n(k) \geq 2l\ T_1 + n - T_l = h) \\
&\leq 2c'_\theta e^{\alpha i} u(i)\gamma^n n^{-3/2}(k+1)e^{\theta k} \sum_{n/2 \leq h \leq n-l} (n-h)^{-3/2} \\
&\leq 2c'_\theta e^{\alpha i} u(i)\gamma^n n^{-3/2}(k+1)e^{\theta k} \sum_{h \geq l} h^{-3/2} \\
&\leq 2.2c'_\theta e^{\alpha i} u(i)\gamma^n n^{-3/2}(k+1)e^{\theta k}/\sqrt{l}, \quad (n \geq n_0).
\end{aligned}$$

Then, using again (7.20),

$$\limsup_{n\to\infty} \mathbb{P}(L_n \geq -i,\ N_n(k) \geq 2l)/\mathbb{P}(L_n \geq -i) \leq 4c'_\theta c_0^{-1}(k+1)e^{\theta k}/\sqrt{l}.$$

Using that $(k+1)e^{\theta k} = o(e^{\theta' k})$ if $\theta' > \theta$, this completes the proof of the first inequality of the lemma for $X \in \mathcal{Z}$. The general case can be proved similarly.

Note that, for every $\theta > \alpha$, when $h \geq n/2$, we can replace (7.28) by

$$\mathbb{P}(L_h \geq -i) \leq 2^{3/2}.c_\theta e^{\theta i} n^{-3/2}\gamma^h, \qquad (i, h, n \in \mathbb{N}).$$

Following the proof above ensures that there exists $c''_\theta > 0$ such for all $i, n, l \in \mathbb{N}$,

$$\mathbb{P}(L_n \geq -i,\ N_n(k) \geq 2l) \leq c''_\theta e^{\theta i}\gamma^n n^{-3/2} e^{\theta k}/\sqrt{l}.$$

Thus, by (7.20), for every $x \geq 0$, there exists $C_x > 0$ such that

$$\mathbb{P}(N_n(k) \geq l \mid L_n \geq -x) \leq 2c''_\theta C_x (k+1)e^{\theta k}/\sqrt{l}, \quad (k, n, l \in \mathbb{N}),$$

which gives the second inequality of the lemma. □

Proof of Corollary 7.4.2. Let $\alpha < 1/2$ and $d > 0$ given by Theorem 7.2.2. Fix $\alpha < \theta < \mu/2 < 1/2$. Choose also $k_0 \in \mathbb{N}$ such that

$$d \sum_{k \geq k_0} e^{(\theta-\mu/2)k} < 1/2.$$

By (8.6.5), for every $x \geq 0$, there exists $D > 0$ such that for every $n \in \mathbb{N}$,

$$\mathbb{P}(N_n(k) \geq e^{\mu k} \mid L_n \geq -x) \leq D e^{(\theta-\mu/2)k}$$

which is summable with respect to k. Thus, by Fatou's lemma,

$$\limsup_{n\to\infty} \sum_{k \geq k_0} \mathbb{P}(N_n(k) \geq e^{\mu k} \mid L_n \geq -x) \leq \sum_{k \geq k_0} \limsup_{n\to\infty} \mathbb{P}(N_n(k) \geq e^{\mu k} \mid L_n \geq -x).$$

By Lemma 7.4.1, this gives, for every $x > 0$,

$$\limsup_{n\to\infty} \sum_{k \geq k_0} \mathbb{P}(N_n(k) \geq e^{\mu k} \mid L_n \geq -x) \leq d \sum_{k \geq k_0} e^{(\theta-\mu/2)k}.$$

Then,

$$\limsup_{n\to\infty} \mathbb{P}\big(\bigcup_{k \geq k_0} \{N_n(k) \geq e^{\mu k}\} \mid L_n \geq -x\big) < 1/2.$$

By Lemma 7.4.1 again, fix $N \in \mathbb{N}$ such that

$$\limsup_{n\to\infty} \mathbb{P}\big(\bigcup_{0 \leq k < k_0} \{N_n(k) \geq N\} \mid L_n \geq -x\big) \leq 1/4.$$

Then

$$\limsup_{n\to\infty} \mathbb{P}\big(\bigcup_{0 \leq k < k_0} \{N_n(k) \geq N\} \bigcup_{k \geq k_0} \{N_k(k) \geq e^{\mu k}\} \mid L_n \geq -x\big) < 3/4.$$

Noting that

$$\sum_{i=0}^{n} \exp(L_n - S_i) \leq \sum_{k=0}^{\infty} N_n(k) e^{-k},$$

this ensures that for every $x \geq 0$,

$$\liminf_{n\to\infty} \mathbb{P}\big(\sum_{i=0}^{n} \exp(L_n - S_i) \leq \beta \mid L_n \geq -x\big) > 1/4,$$

with $\beta := \sum_{0 \leq k < k_0} N e^{-k+1} + \sum_{k \geq k_0} e^{\mu k} e^{-k+1}$. This gives the result. $\square$

Chapter 8

Kimmel's branching model for cell division with parasite infection

8.1 Introduction

We consider the following model for cell division with parasite infection. Unless otherwise specified, we start with a single cell infected with a single parasite. At each generation, each parasite multiplies independently, each cell divides into two daughter cells and the offspring of each parasite is shared independently into the two daughter cells. It is convenient to distinguish a first daughter cell called 0 and a second one called 1 and to write $Z^{(0)} + Z^{(1)}$ the number of offspring of a parasite, $Z^{(0)}$ of which go into the first daughter cell and $Z^{(1)}$ of which into the second one. The symmetric sharing is the case when $(Z^{(0)}, Z^{(1)}) \overset{d}{=} (Z^{(1)}, Z^{(0)})$. Even in that case, the sharing of parasites can be unequal (for example when $\mathbb{P}(Z^{(0)}Z^{(1)} = 0) = 1$).

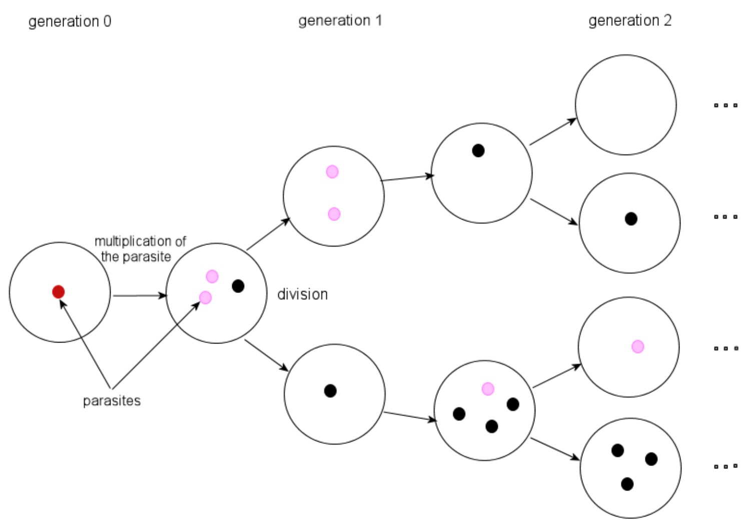
We denote by $\mathbb{T}$ the binary genealogical tree of the cell population, by $\mathbb{G}_n$ (resp. $\mathbb{G}_n^*$) the set of cells at generation n (resp. the set of contaminated cells at generation n) and by $Z_{\mathbf{i}}$ the number of parasites of cell $\mathbf{i} \in \mathbb{T}$, i.e.

$$\mathbb{G}_n := \{0, 1\}^n, \qquad \mathbb{G}_n^* := \{\mathbf{i} \in \mathbb{G}_n : Z_{\mathbf{i}} > 0\}, \qquad \mathbb{T} := \cup_{n \in \mathbb{N}} \mathbb{G}_n.$$

For every cell $\mathbf{i} \in \mathbb{T}$, conditionally on $Z_{\mathbf{i}} = x$, the numbers of parasites $(Z_{\mathbf{i}0}, Z_{\mathbf{i}1})$ of its two daughter cells is given by

$$\sum_{k=1}^{x} (Z_k^{(0)}(\mathbf{i}), Z_k^{(1)}(\mathbf{i})),$$

where $(Z_k^{(0)}(\mathbf{i}), Z_k^{(1)}(\mathbf{i}))_{\mathbf{i} \in \mathbb{T}, k \geq 1}$ is an iid sequence distributed as $(Z^{(0)}, Z^{(1)})$.



This is a discrete version of the model introduced by M. Kimmel in [59]. In particular, it contains the following model with binomial repartition of parasites. Let Z be a random variable in $\mathbb{N}$ and $p \in [0,1]$. At each generation, every parasite multiplies independently with the same reproduction law Z. When the cells divides, every parasite chooses independently the first daughter cell with probability p (and the second one with probability $1-p$). It contains also the case when every parasite gives birth to a random cluster of parasites of size Z which goes to the first cell with probability p (and to the second one with probability $1-p$).

We introduce for $a \in \{0,1\}$

$$m_a := \mathbb{E}(Z^{(a)}), \qquad \forall\ s \geq 0,\ f_a(s) := \mathbb{E}(s^{Z^{(a)}}). \tag{8.1}$$

We assume $0 < m_0 < \infty$, $0 < m_1 < \infty$ and to avoid trivial cases, we require

$$\mathbb{P}((Z^{(0)}, Z^{(1)}) = (1,1)) < 1, \qquad \mathbb{P}((Z^{(0)}, Z^{(1)}) \in \{(1,0),(0,1)\}) < 1. \tag{8.2}$$

This model is a Markov chain indexed by a tree. This subject has been studied in the literature (see e.g. [9, 18]) in the symmetric independent case. In this case, for every $(\mathbf{i}, k) \in \mathbb{T} \times \mathbb{N}$, we have

$$\mathbb{P}((Z_{\mathbf{i}0}, Z_{\mathbf{i}1}) = (k_0, k_1) \mid Z_{\mathbf{i}} = k) = \mathbb{P}(Z_{\mathbf{i}0} = k_0 \mid Z_{\mathbf{i}} = k)\mathbb{P}(Z_{\mathbf{i}0} = k_1 \mid Z_{\mathbf{i}} = k)$$

which require that $Z^{(0)}$ and $Z^{(1)}$ are iid in this model. Guyon [47] studies a Markov chain indexed by a binary tree where asymmetry and dependence are allowed and limit theorems are proved. But the case where his results apply is degenerate (this is the case $m_0 m_1 \leq 1$ and the limit of the number of parasites in a random cell line is zero). Moreover adapting his arguments for the theorems stated here appears to be cumbersome (see the remark in Section 5.2 for details). In the same vein, we refer to [35, 88] (cellular aging).

The total population of parasites at generation n, which we denote by $\mathcal{Z}_n$, is a Bienaymé Galton Watson process (BGW) with reproduction law $Z^{(0)} + Z^{(1)}$. We call Ext (resp Ext^c) the event Extinction of the parasites (resp Non extinction of the parasites),

$$\mathcal{Z}_n = \sum_{\mathbf{i} \in \mathbb{G}_n} Z_{\mathbf{i}}, \quad \text{Ext} = \{\exists n \in \mathbb{N} : \mathcal{Z}_n = 0\}, \quad \text{Ext}^c = \{\forall n \in \mathbb{N} : \mathcal{Z}_n > 0\}. \tag{8.3}$$

Another process that appears naturally is the number of parasites in a random cell line. More precisely, let $(a_i)_{i \in \mathbb{N}}$ be an iid sequence independent of $(Z_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$ such that

$$\mathbb{P}(a_1 = 0) = \mathbb{P}(a_1 = 1) = 1/2. \tag{8.4}$$

Then $(Z_n)_{n \in \mathbb{N}} = (Z_{(a_1, a_2, ..a_n)})_{n \in \mathbb{N}}$ is a Branching Process in Random Environment (BPRE).

The first question we answer here arose from observations made by M. de Paepe, G. Paul and F. Taddei at TaMaRa's Laboratory (Hôpital Necker, Paris). They have infected the bacteria *E. Coli* with a parasite (lysogen bacteriophage M13). A fluorescent marker allows them to see the level of contamination of cells. They observed that a very contaminated cell often gives birth to a very contaminated cell which dies fast and to a much less contaminated cell whose descendance may survive. So cells tend to share unequally their parasites when they divide so that there are lots of healthy cells. This is a little surprising since one could think that cells share equally all their biological content (including parasites). In Section 3, we prove that if $m_0 m_1 \leq 1$, the organism recovers a.s. (meaning that the number of infected cells becomes negligible compared to the number of cells when $n \to \infty$). Otherwise the organism recovers iff parasites die out (and the probability is less than 1).

In Section 4, we consider the tree of contaminated cells. We denote by $\partial\mathbb{T}$ the boundary of the cell tree $\mathbb{T}$ and by $\partial\mathbb{T}^*$ the infinite lines of contaminated cells, that is,

$$\partial\mathbb{T} = \{0, 1\}^{\mathbb{N}}, \qquad \partial\mathbb{T}^* = \{\mathbf{i} \in \partial\mathbb{T} : \forall n \in \mathbb{N},\ Z_{\mathbf{i}|n} \neq 0\}.$$

We shall prove that the contaminated cells are not concentrated in a cell line. Note that if $m_0 + m_1 > 1$, conditionally on Ext^c, $\partial\mathbb{T}^* \neq \varnothing$ since at each generation, one can choose a daughter cell whose parasite descendance does not become extinct.

The rest of the work is devoted to the convergence of the number of contaminated cells in generation n and the convergence of proportions of contaminated cells with a given number of parasites (Section 5). These asymptotics depend on

(m_0, m_1) and we distinguish five different cases which come from the behavior of the BGW process $\mathcal{Z}_n$ and the BPRE Z_n (Section 2).

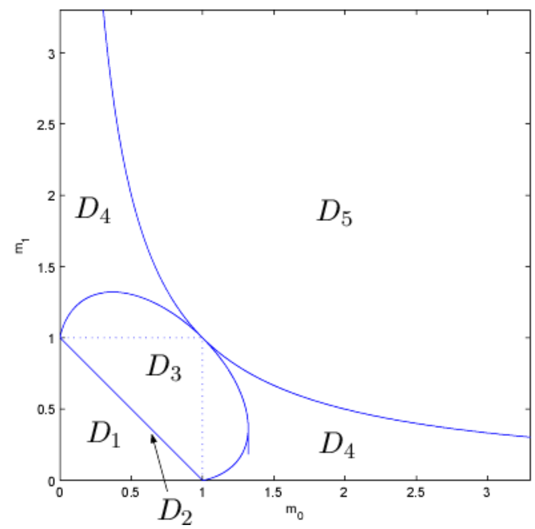
$$D_1 = \{(m_0, m_1) : m_0 + m_1 < 1\}$$

$$D_2 = \{(m_0, m_1) : m_0 + m_1 = 1\}$$

$$D_3 = \{(m_0, m_1) : m_0 + m_1 > 1, \quad m_0 \log(m_0) + m_1 \log(m_1) < 0\}$$

$$D_4 = \{(m_0, m_1) : m_0 m_1 \le 1, \quad m_0 \log(m_0) + m_1 \log(m_1) \ge 0\}$$

$$D_5 = \{(m_0, m_1) : m_0 m_1 > 1\}$$



If $(m_0, m_1) \in D_5$, the contaminated cells become largely infected (Theorem 8.5.1). The main two results correspond to cases $(m_0, m_1) \in D_3$ and $(m_0, m_1) \in D_1$ and are given by the following two theorems.

Theorem 5.2. *If $(m_0, m_1) \in D_3$, conditionally on Ext^c, the following convergence holds in probability for every $k \in \mathbb{N}$,*

$$\#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} = k\} / \#\mathbb{G}_n^* \stackrel{n \to \infty}{\longrightarrow} \mathbb{P}(\Upsilon = k),$$

where Υ is the Yaglom quasistationary distribution of the BPRE $(Z_n)_{n \in \mathbb{N}}$ (see [8, 43]). Note that the limit is deterministic and depends solely on the marginal laws of $(Z^{(0)}, Z^{(1)})$ (see Proposition 8.2.2). This gives then a way to compute Υ as a deterministic limit, although it is defined by conditioning on a vanishing event. Kimmel [59] considers the symmetric case $((Z^{(0)}, Z^{(1)}) \stackrel{d}{=} (Z^{(1)}, Z^{(0)}))$ with $m_0 = m_1 < 1 < m_0 + m_1$ in a continuous analogue of this model (cells divide after an exponential time). The counterpart of his result in the discrete case is easy to prove (see (8.20)) and makes a first link with Υ.

$$\lim_{n \to \infty} \mathbb{E}(\#\{\mathbf{i} \in \mathbb{G}_n : Z_{\mathbf{i}} = k\}) / \mathbb{E}(\#\mathbb{G}_n^*) = \mathbb{P}(\Upsilon = k).$$

Moreover the proportions of contaminated cells on the boundary of the tree whose ancestors at generation n have a given number of parasites converge to the size-biased distribution of Υ letting $n \to \infty$ (Corollary 8.5.4). This gives a pathwise interpretation that the limit of the Q-process associated to Z_n (see [1, 8]) is the size-biased quasistationary distribution.

Theorem 5.7. *If $(m_0, m_1) \in D_1$, $(\#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} = k\})_{k\in\mathbb{N}}$ conditioned on $\mathcal{Z}_n > 0$ converges in distribution as $n \to \infty$ to a finite random sequence $(N_k)_{k\in\mathbb{N}}$.*

We obtain a similar result in the case $(m_0, m_1) \in D_2$ (Theorem 8.5.5) and we get the following asymptotics (Theorem 8.3.1 and Corollaries 8.5.3, 8.5.6, 8.5.8).

- ⋆ If $(m_0, m_1) \in D_3$ (resp. D_5), then conditionally on Ext^c, $\#\mathbb{G}_n^*/(m_0 + m_1)^n$ (resp $\#\mathbb{G}_n^*/2^n$) converges in probability to a finite positive r.v.
- ⋆ If $(m_0, m_1) \in D_1$ (resp. D_2), then $\#\mathbb{G}_n^*$ (resp $\#\mathbb{G}_n^*/n$) conditioned by $\#\mathbb{G}_n^* > 0$ converges in distribution to a finite positive r.v.

In the case $(m_0, m_1) \in D_4$, we get only some estimates of the asymptotic of $\#\mathbb{G}_n^*$ which are different from those which hold in the other domains. Our conjecture is that $\#\mathbb{G}_n^*$ has also a deterministic asymptotic, which depends on three subdomains (the interior of D_4 and its boundaries). As a perspective, we are also interested in determining which types of convergences hold in D_4 for the proportions of contaminated cells with a given number of parasites (see Section 5.5).

Moreover we wonder if the convergences stated above hold a.s. and if they extend to the continuous case and complement the results of Kimmel. Finally, in a work in progress with Julien Beresticky and Amaury Lambert, we aim at determining the localizations of contaminated cells and the presence of cells filled-in by parasites on the boundary of the tree (branching measure and multifractal analysis).

8.2 Preliminaries

In this section, we give some useful results about the two processes introduced above. First define :

$$m := \frac{1}{2}(m_0 + m_1). \tag{8.5}$$

We use the classical notation, where for every $\mathbf{i} = (\alpha_1, .., \alpha_n) \in \mathbb{G}_n$,

$$|\mathbf{i}| = n, \qquad \mathbf{i}|k = (\alpha_1, .., \alpha_k) \text{ for every } k \leq n, \qquad \mathbf{j} < \mathbf{i} \text{ if } \exists k < n \ : \ \mathbf{i}|k = \mathbf{j}.$$

8.2.1 Results on the BGW process $(\mathcal{Z}_n)_{n\in\mathbb{N}}$

The results stated hereafter are well known and can be found in [8]. First the probability of extinction of the parasites satisfies

$$\mathbb{P}(\mathrm{Ext}) = \inf\{s \in [0,1] : \mathbb{E}\big(s^{Z^{(0)}+Z^{(1)}}\big) = s\} \quad ; \quad \mathbb{P}(\mathrm{Ext}) = 1 \ \text{ iff } m_0 + m_1 \leq 1/2.$$

From now, we assume

$$\check{m} := \mathbb{E}((Z^{(0)} + Z^{(1)})\log^+(Z^{(0)} + Z^{(1)})) < \infty.$$

Then there exists a random variable W such that

$$\frac{\mathcal{Z}_n}{(m_0 + m_1)^n} \overset{n\to\infty}{\longrightarrow} W, \qquad \mathbb{P}(W = 0) = \mathbb{P}(\text{Ext}), \qquad \mathbb{E}(W) = 1. \tag{8.6}$$

In the case $m_0 + m_1 < 1$, there exists $b > 0$ such that $\mathbb{P}(\mathcal{Z}_n > 0) \overset{n\to\infty}{\sim} b(m_0 + m_1)^n$. Then, there exists $U > 0$ such that

$$\mathbb{P}(\mathcal{Z}_n > 0) \geq U(m_0 + m_1)^n. \tag{8.7}$$

Moreover $(\mathcal{Z}_n)_{n\in\mathbb{N}}$ conditioned to be non zero converges to a variable called the Yaglom quasistationary distribution and we set

$$\mathcal{B}(s) := \lim_{n\to\infty} \mathbb{E}(s^{\mathcal{Z}_n} \mid \mathcal{Z}_n > 0). \tag{8.8}$$

We consider then $\mathcal{B}_{n,k}(s) := \mathbb{E}(s^{\mathcal{Z}_n} \mid \mathcal{Z}_{n+k} > 0)$ which satisfies

$$\lim_{n\to\infty} \mathcal{B}_{n,k}(s) = \frac{\mathcal{B}(s) - \mathcal{B}(sf_k(0))}{1 - \mathcal{B}(f_k(0))}. \tag{8.9}$$

Moreover $\mathcal{B}$ is differentiable at 1 (Lemma 1 on page 44 in [8]) and we get

$$\lim_{k\to\infty} \lim_{n\to\infty} \mathcal{B}_{n,k}(s) = \frac{s\mathcal{B}'(s)}{\mathbf{B}_0'(1)}. \tag{8.10}$$

This is the probability generating function of the size-biased Yaglom quasistationary distribution, which is also the stationary distribution of the Q-process.

Finally if $\hat{m} := \mathbb{E}((Z^{(0)} + Z^{(1)})((Z^{(0)} + Z^{(1)}) - 1)) < \infty$ and $2m \neq 1$, then

$$\mathbb{E}(\mathcal{Z}_n(\mathcal{Z}_n - 1)) = \hat{m}(2m)^n \frac{(2m)^n - 1}{(2m)^2 - 2m}. \tag{8.11}$$

8.2.2 Properties of the BPRE $(Z_n)_{n\in\mathbb{N}}$

Recall that $(Z_n)_{n\in\mathbb{N}}$ is the population of parasites in a uniform random cell line. Then $(Z_n)_{n\in\mathbb{N}}$ is a BPRE with two equiprobable environments. More precisely, for each $n \in \mathbb{N}$, conditionally on $a_n = a$ with $a \in \{0, 1\}$ (see (9.2)), all parasites behave independently of one another and each of them gives birth to $Z^{(a)}$ children. The size of the population at generation 0 is denoted by k and we note

$\mathbb{P}_k$ the associated probability. Unless otherwise mentioned, the initial state is equal to 1. For the general theory, see e.g. [31, 43, 46, 87]. In the case $Z^{(0)} \stackrel{d}{=} Z^{(1)}$, $(Z_n)_{n\in\mathbb{N}}$ is a BGW with reproduction law $Z^{(0)}$.

For $\mathbf{i} = (\alpha_1, .., \alpha_n) \in \mathbb{G}_n$, we define

$$f_{\mathbf{i}} := f_{\alpha_1} \circ .. \circ f_{\alpha_n}, \qquad m_{\mathbf{i}} = \prod_{i=1}^{n} m_{\alpha_i},$$

and for all $(n,k) \in \mathbb{N} \times \mathbb{N}^*$ and $\mathbf{i} \in \mathbb{G}_n$,

$$\mathbb{E}_k\big(s^{Z_n} \mid (a_1, .., a_n) = \mathbf{i}\big) = f_{\mathbf{i}}(s)^k.$$

Then for all $(n,k) \in \mathbb{N} \times \mathbb{N}^*$ and $s \in [0,1]$,

$$\mathbb{E}_k\big(s^{Z_n}\big) = 2^{-n} \sum_{\mathbf{i}\in\mathbb{G}_n} f_{\mathbf{i}}(s)^k. \tag{8.12}$$

First, for every $n \in \mathbb{N}$, $\mathbb{E}(Z_{n+1} \mid Z_n) = mZ_n$ and $\mathbb{E}(Z_n) = m^n$.

Moreover, as $(\mathbb{P}(Z_n = 0))_{n\in\mathbb{N}}$ is an increasing sequence, it converges to the probability of extinction p of the process. Recalling (8.1), we have the following result (see [87] or [6] or Chapter 6).

Proposition 8.2.1. *If $m_0m_1 \leq 1$, then $p = 1$.Otherwise $p < 1$.*

In the subcritical case ($m_0m_1 < 1$), the process Z_n conditioned to be non zero which is denoted by Z_n^* converges weakly to the Yaglom quasistationary distribution which is denoted by Υ (see Theorem 1.1 in [43] or Chapter 6). That is,

$$\forall s \in [0,1], \quad \mathbb{E}\big(s^{Z_n} | Z_n > 0\big) \stackrel{n\to\infty}{\longrightarrow} \mathbb{E}(s^{\Upsilon}) = G(s).$$

In the subcritical case, the asymptotics of $(\mathbb{P}(Z_n > 0))_{n\in\mathbb{N}}$ when n is large depends on the sign of $m_0\log(m_0) + m_1\log(m_1)$ (see [43] or Chapter 6). Now, we introduce the following condition

$$m_0\log(m_0) + m_1\log(m_1) < 0 \quad ; \quad \mathbb{E}(Z_a\log^+(Z_a)) < \infty. \tag{8.13}$$

Recall that in this case, we say that Z_n is strongly subcritical and there exists $c > 0$ such that as n tends to ∞ (by 7.2 or Theorem 1.1 in [43]),

$$\mathbb{P}(Z_n > 0) \sim cm^n. \tag{8.14}$$

Moreover, in that case, Theorem 7.2.7 in the previous chapter ensures that Υ is characterized by

Proposition 8.2.2. *G is the unique probability generating function which satisfies*

$$G(0) = 0, \quad G'(1) < \infty, \quad \frac{G(f_0(s)) + G(f_1(s))}{2} = mG(s) + (1-m). \tag{8.15}$$

In the subcritical case ($m_0m_1 \leq 1$), if $m_0\log(m_0) + m_1\log(m_1) > 0$ (resp. $m_0\log(m_0) + m_1\log(m_1) = 0$)), recall that we say that Z_n is weakly subcritical (resp. intermediate subcritical) and we have $\mathbb{P}(Z_n > 0) \sim c'n^{-3/2}\gamma^n$ (resp. $\mathbb{P}(Z_n > 0) \sim c''n^{-1/2}m^n$) for some $\gamma < m, c' > 0, c'' > 0$ (see [43] for details or Chapter 6).

8.3 Probability of recovery

We say that the organism recovers if the number of contaminated cells becomes negligible compared to the number of cells when $n \to \infty$. We determine here the probability of this event. Actually if this probability is not equal to 1, then the parasites must die out for the organism to recover.

Theorem 8.3.1. *There exists a random variable $L \in [0, 1]$ such that*

$$\#\mathbb{G}_n^*/2^n \overset{n\to\infty}{\longrightarrow} L.$$

If $m_0m_1 \leq 1$ then $\mathbb{P}(L = 0) = 1$.
Otherwise $\mathbb{P}(L = 0) < 1$ and $\{L = 0\} = \mathrm{Ext}$.

Remark 11. In the case $m_0 + m_1 > 1$ and $m_0m_1 \leq 1$, the population of parasites may explode although the organism recovers.
This theorem states how unequal the sharing of parasites must be for the organism to recover. More precisely, let $m_0 = \alpha M,\ m_1 = (1-\alpha)M$ where $M > 0$ is the parasite growth rate. Then the organism recovers a.s. iff

$$M \leq 2 \qquad \text{or} \qquad \alpha \notin \,](1-\sqrt{1-4/M^2})/2, (1+\sqrt{1-4/M^2})/2[\quad (M > 2).$$

Note that for all $n \in \mathbb{N}$,

$$\mathbb{E}\big(\frac{\#\mathbb{G}_n^*}{2^n}\big) = \frac{\mathbb{E}(\sum_{i\in\mathbb{G}_n} \mathbb{1}_{Z_\mathbf{i}>0})}{2^n} = \mathbb{P}(Z_n > 0).$$

Recalling that p is the probability of extinction of $(Z_n)_{n\in\mathbb{N}}$,

$$\forall n \in \mathbb{N}, \qquad \mathbb{E}\big(\frac{\#\mathbb{G}_n^*}{2^n}\big) = \mathbb{P}(Z_n > 0) \overset{n\to\infty}{\longrightarrow} 1 - p. \tag{8.16}$$

The last equality gives also the asymptotic of $\mathbb{E}(\#\mathbb{G}_n^*)$ as $n \to \infty$ in the case $m_0m_1 < 1$ (see Section 2.2 for the asymptotic of $\mathbb{P}(Z_n > 0)$, which depends on the sign of $m_0\log(m_0) + m_1\log(m_1)$) and in the case $m_0m_1 = 1$ (see [**?**, 63]).

Proof of Theorem 8.3.1. As $\#\mathbb{G}_n^*/2^n$ decreases as n increases, it converges as $n \to \infty$.

Monotone convergence of $\#\mathbb{G}_n^*/2^n$ to L as $n \to \infty$ and (8.16) ensure that $\mathbb{E}(L) = 1 - p$. Using Proposition 8.2.1, we get $\mathbb{P}(L = 0) = 1$ iff $m_0 m_1 \leq 1$.

Obviously $\{L = 0\} \supset \mathrm{Ext}$. Denote by $\mathcal{P}(n)$ the set of parasites at generation n and for every $\mathbf{p} \in \mathcal{P}(n)$, denote by $N_k(\mathbf{p})$ the number of cells at generation $n + k$ which contain at least a parasite whose ancestor is $\mathbf{p}$. Then, for every $n \in \mathbb{N}$,

$$\{L = 0\} = \bigcap_{\mathbf{p} \in \mathcal{P}(n)} \Big\{ \frac{N_k(\mathbf{p})}{2^k} \stackrel{k \to \infty}{\longrightarrow} 0 \Big\}.$$

As $T_n := \inf\{k \geq 0 : \mathcal{Z}_k \geq n\}$ is a stopping time with respect to the natural filtration of $(Z_{\mathbf{i}})_{|\mathbf{i}| \leq n}$, strong Markov property gives

$$\mathbb{P}(L = 0) \leq \mathbb{P}(T_n < \infty)\mathbb{P}(L = 0)^n + \mathbb{P}(T_n = \infty)$$

If $\mathbb{P}(L = 0) < 1$, letting $n \to \infty$ gives

$$\mathbb{P}(L = 0) \leq \lim_{n \to \infty} \mathbb{P}(T_n = \infty) = \mathbb{P}(\mathcal{Z}_n \text{ is bounded}) = \mathbb{P}(\mathrm{Ext})$$

since $\mathcal{Z}_n$ is a BGW. This completes the proof. One can also use a coupling argument : the number of contaminated cells starting with one single cell with n parasites is less than the number of contaminated cells starting from n cells with one single parasite. $\square$

8.4 Tree of contaminated cells

Here, we prove that contaminated cells are not concentrated in a cell line. If $m_0 + m_1 \leq 1$, contaminated cells die out but conditionally on the survival of parasites at generation n, the number of leaves of the tree of contaminated cells tends to ∞ as $n \to \infty$. The proof of this result will also ensure that, if $m_0 + m_1 > 1$, the number of contaminated cells tends to ∞ provided that they do not die out.

Theorem 8.4.1. *If $m_0 + m_1 \leq 1$, $\#\{\mathbf{i} \in \mathbb{T} : Z_{\mathbf{i}} \neq 0, Z_{\mathbf{i}0} = 0, Z_{\mathbf{i}1} = 0\}$ conditioned by $\#\mathbb{G}_n^* > 0$ converges in probability as $n \to \infty$ to ∞.*
If $m_0 + m_1 > 1$, conditionally on Ext^c, $\#\mathbb{G}_n^* \stackrel{n \to \infty}{\longrightarrow} \infty$ a.s.

Remark 12. In the conditions of the theorem, $\#\mathbb{G}_n^*$ (resp. the number of leaves) grows at least linearly with respect to n (see Section 5 for further results). In the case $m_0 + m_1 \leq 1$, conditionally on $\#\mathbb{G}_n^* > 0$, the tree of contaminated cells is a spine with finite subtrees, as for BGW conditioned to survive (see [42, 69]).

We need two lemmas for the proof. First we prove that the ancestor of a contaminated cell has given birth to two contaminated cells with a probability bounded from below. We have to distinguish the case where $\mathbb{P}(Z^{(0)}Z^{(1)}=0)=1$, since in that case a cell must contain at least two parasites so that it can give birth to two contaminated cells.

Lemma 8.4.2. *There exists $\alpha>0$ such that for all $N\in\mathbb{N}$, $\mathbf{i}\in\mathbb{G}_N$, $n<N$ and $k\geq 2$,*

$$\mathbb{P}(Z_{\mathbf{j}0}\neq 0, Z_{\mathbf{j}1}\neq 0 \ \mid\ Z_{\mathbf{j}}=k, Z_{\mathbf{i}}>0)\geq \alpha$$

denoting $\mathbf{j}=\mathbf{i}\mid n$. If $\mathbb{P}(Z^{(0)}Z^{(1)}=0)\neq 1$, this result also holds for $k=1$.

Proof. We consider first the case $\mathbb{P}(Z^{(0)}Z^{(1)}=0)\neq 1$ and we choose $(k_0,k_1)\in\mathbb{N}^{*2}$ such that $\mathbb{P}((Z^{(0)},Z^{(1)})=(k_0,k_1))>0$. For every $k\in\mathbb{N}^*$, we have

$$\mathbb{P}(Z_{\mathbf{j}0}\neq 0, Z_{\mathbf{j}1}\neq 0 \ \mid\ Z_{\mathbf{j}}=k, Z_{\mathbf{i}}>0)\geq \mathbb{P}(Z_{\mathbf{j}0}\neq 0, Z_{\mathbf{j}1}\neq 0 \ \mid\ Z_{\mathbf{j}}=1, Z_{\mathbf{i}}>0)$$

Moreover, as the function $\mathbb{R}_+^*\ni u\mapsto \frac{1-e^{-u}}{u}$ decreases, we have for all $y,x>0$ and $p\in[0,1[$,

$$\frac{1-p^x}{1-p^y}\geq \frac{x}{\max\{y,x\}}. \tag{8.17}$$

Let $a\in\{0,1\}$ and $\mathbf{k}$ such that $\mathbf{i}=\mathbf{j}a\mathbf{k}$. Then for all $(k_0',k_1')\in\mathbb{N}^2-(0,0)$,

$$\begin{aligned}
&\frac{\mathbb{P}(Z_{\mathbf{j}0}=k_0, Z_{\mathbf{j}1}=k_1 \ \mid\ Z_{\mathbf{j}}=1, Z_{\mathbf{i}}>0)}{\mathbb{P}(Z_{\mathbf{j}0}=k_0', Z_{\mathbf{j}1}=k_1' \ \mid\ Z_{\mathbf{j}}=1, Z_{\mathbf{i}}>0)}\\
&\qquad = \frac{\mathbb{P}(Z^{(0)}=k_0, Z^{(1)}=k_1 \ \mid\ Z_{a\mathbf{k}}>0)}{\mathbb{P}(Z^{(0)}=k_0', Z^{(1)}=k_1' \ \mid\ Z_{a\mathbf{k}}>0)}\\
&\qquad = \frac{\mathbb{P}(Z_{a\mathbf{k}}>0 \ \mid\ Z^{(0)}=k_0, Z^{(1)}=k_1)}{\mathbb{P}(Z_{a\mathbf{k}}>0 \ \mid\ Z^{(0)}=k_0', Z^{(1)}=k_1')}\frac{\mathbb{P}(Z^{(0)}=k_0, Z^{(1)}=k_1)}{\mathbb{P}(Z^{(0)}=k_0', Z^{(1)}=k_1')}\\
&\qquad = \frac{1-\mathbb{P}(Z_{\mathbf{k}}=0)^{k_a}}{1-\mathbb{P}(Z_{\mathbf{k}}=0)^{k_a'}}\frac{\mathbb{P}((Z^{(0)},Z^{(1)})=(k_0,k_1))}{\mathbb{P}((Z^{(0)},Z^{(1)})=(k_0',k_1'))}\\
&\qquad \geq \frac{\min\{k_0,k_1\}}{k_0+k_1+k_0'+k_1'}\frac{\mathbb{P}((Z^{(0)},Z^{(1)})=(k_0,k_1))}{\mathbb{P}((Z^{(0)},Z^{(1)})=(k_0',k_1'))} \quad \text{using (8.17)}
\end{aligned}$$

Cross product and sum over (k_0',k_1') give

$$\begin{aligned}
\big[\mathbb{E}(Z^{(0)}+Z^{(1)})+k_0+k_1\big]&\mathbb{P}(Z_{\mathbf{j}0}=k_0, Z_{\mathbf{j}1}=k_1 \mid Z_{\mathbf{j}}=1, Z_{\mathbf{i}}>0)\\
&\geq \min\{k_0,k_1\}\mathbb{P}((Z^{(0)},Z^{(1)})=(k_0,k_1)).
\end{aligned}$$

This gives the result since $\mathbb{P}(Z_{\mathbf{j}0}=k_0, Z_{\mathbf{j}1}=k_1 \mid Z_{\mathbf{j}}=1, Z_{\mathbf{i}}>0)\geq\alpha$ with

$$\alpha=\frac{\min\{k_0,k_1\}\mathbb{P}((Z^{(0)},Z^{(1)})=(k_0,k_1))}{\mathbb{E}(Z^{(0)}+Z^{(1)})+k_0+k_1}>0.$$

In the case $\mathbb{P}(Z^{(0)}Z^{(1)}=0)=1$, we choose $(k_0,k_1)\in\mathbb{N}^{*2}$ such that $\mathbb{P}_2((Z_0,Z_1)=(k_0,k_1))>0$ (using (8.2)). We make then the same proof as above with $Z_{\mathbf{j}}=2$ and

$$\alpha=\frac{\min\{k_0,k_1\}\mathbb{P}_2((Z_0,Z_1)=(k_0,k_1))}{\mathbb{E}_2(Z_0+Z_1)+k_0+k_1},$$

so that the result follows as previously. □

Thus if $\mathbb{P}(Z^{(0)}Z^{(1)}=0)=1$, we need to prove that there are many cells with more than two parasites in a contaminated cell line.

Lemma 8.4.3. *If* $\beta := \mathbb{P}(Z^{(0)} \geq 2 \text{ or } Z^{(1)} \geq 2) > 0$ *then*

$$\inf_{\mathbf{i}\in\mathbb{G}_n} \mathbb{P}(\#\{\mathbf{j} < \mathbf{i} : Z_{\mathbf{j}0} \geq 2 \text{ or } Z_{\mathbf{j}1} \geq 2\} \geq \beta n/2 \mid Z_{\mathbf{i}} > 0) \overset{n\to\infty}{\longrightarrow} 1.$$

Proof. For all $\mathbf{i} \in \mathbb{G}_n$ and $\mathbf{j} < \mathbf{i}$, let $\mathbf{k}$ such that $\mathbf{i} = \mathbf{jk}$, then for every $\alpha > 0$,

$$\mathbb{P}(Z_{\mathbf{j}0} \geq 2 \text{ or } Z_{\mathbf{j}1} \geq 2 \mid Z_{\mathbf{j}} = \alpha,\ Z_{\mathbf{i}} > 0) \geq \mathbb{P}(Z_0 \geq 2 \text{ or } Z_1 \geq 2 \mid Z_{\mathbf{k}} > 0) \geq \beta$$

Then conditionally on $Z_{\mathbf{i}} > 0$, $\#\{\mathbf{j} < \mathbf{i} : Z_{\mathbf{j}0} \geq 2 \text{ or } Z_{\mathbf{j}1} \geq 2\} \geq \sum_{k=0}^{n} \beta_k$, where $(\beta_k)_{1\leq k\leq n}$ are iid and distributed as a Bernoulli(β). Conclude with the law of large numbers. $\square$

Proof of Theorem 8.4.1. We consider first the case when $m_0 + m_1 > 1$, work conditionally on Ext^c and choose $\mathbf{i} \in \delta\mathbb{T}^*$.

If $\mathbb{P}(Z^{(0)}Z^{(1)} = 0) \neq 1$, Lemma 8.4.2 (with $k = 1$) entails that a.s. under $\mathbb{P}(.|Z_{\mathbf{i}} > 0)$,

$$\#\{\mathbf{j} < \mathbf{i} : Z_{\mathbf{j}0} > 0, Z_{\mathbf{j}1} > 0\} = \infty.$$

Using the branching property and the fact that the probability of non-extinction of parasites is positive ensures that $\#\mathbb{G}_n^* \overset{n\to\infty}{\longrightarrow} \infty$ a.s.

If $\mathbb{P}(Z^{(0)}Z^{(1)} = 0) = 1$ then $\mathbb{P}(Z^{(0)} \geq 2 \text{ or } Z^{(1)} \geq 2) > 0$ and by Lemma 8.4.3, we have a.s. on $\mathbb{P}(.|Z_{\mathbf{i}} > 0)$,

$$\#\{\mathbf{j} < \mathbf{i} : Z_{\mathbf{j}0} \geq 2 \text{ or } Z_{\mathbf{j}1} \geq 2\} = \infty.$$

Using as above Lemma 8.4.2 (with $k = 2$) and the fact that the probability of non-extinction of parasites is positive ensures that $\#\mathbb{G}_n^* \overset{n\to\infty}{\longrightarrow} \infty$ a.s.

We consider now the case when $m_0 + m_1 \leq 1$ and work conditionally on $\mathbf{i} = (\alpha_1, .., \alpha_n) \in \mathbb{G}_n^*$. We denote $\mathbf{i}_j := (\alpha_1, .., \alpha_{j-1}, 1-\alpha_j)$ for $1 \leq j \leq n$.

If $\mathbb{P}(Z^{(0)}Z^{(1)} = 0) \neq 1$, Lemma 8.4.2 entails that

$$\forall 1 \leq j \leq n,\ k \geq 1, \quad \mathbb{P}(Z_{\mathbf{i}_j} > 0 \mid Z_{\mathbf{i}|j-1} = k,\ Z_{\mathbf{i}} > 0) \geq \alpha. \tag{8.18}$$

Moreover if $Z_{\mathbf{i}_j} > 0$, then the tree of contaminated cells rooted in $\mathbf{i}_j$ dies out and so has at least one leaf. So by the branching property, the number of leaves converges in probability to infinity as n tends to infinity.

If $\mathbb{P}(Z^{(0)}Z^{(1)} = 0) = 1$, (8.18) holds for $k \geq 2$ and Lemma 8.4.3 allows to conclude similarly in this case. □

8.5 Proportion of contaminated cells with a given number of parasites

We determine here the asymptotics of the number of contaminated cells and the proportion F_k of cells with k parasites, defined as

$$F_k(n) := \frac{\#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} = k\}}{\#\mathbb{G}_n^*} \qquad (k \in \mathbb{N}^*).$$

In that view, we introduce the Banach space $l^1(\mathbb{N})$ and the subset of frequencies $\mathbb{S}^1(\mathbb{N})$ which we endow with the norm $\| \, . \, \|_1$ defined by :

$$l^1(\mathbb{N}) := \{(x_i)_{i\in\mathbb{N}} : \sum_{i=0}^{\infty} |x_i| < \infty\}, \qquad \| (x_i)_{i\in\mathbb{N}} \|_1 = \sum_{i=0}^{\infty} |x_i|,$$

$$\mathbb{S}^1(\mathbb{N}) := \{(f_i)_{i\in\mathbb{N}} : \forall\, i \in \mathbb{N},\ f_i \in \mathbb{R}^+,\ \sum_{i=0}^{\infty} f_i = 1\}.$$

We shall work conditionally on Ext^c or $\mathcal{Z}_n > 0$ and introduce

$$\mathbb{P}^* := \mathbb{P}(\ .\ \mid \mathrm{Ext}^c), \qquad \mathbb{P}^n := \mathbb{P}(\ .\ \mid \mathcal{Z}_n > 0). \tag{8.19}$$

The asymptotics of the proportions depend naturally on the distribution of $(Z^{(0)}, Z^{(1)})$ and we determine five different behaviors according to the bivariate value of (m_0, m_1).
The proofs of the convergences use the asymptotic distribution of the number of parasites of a typical contaminated cell at generation n, which is equal to $\mathbb{P}^n(Z_{\mathbf{U}_n} \in .)$, where $\mathbf{U}_n$ is a uniform random variable in $\mathbb{G}_n^*$ independent of $(Z_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}^*}$. This distribution is different from the distribution of Z_n^*, that is the number of parasites of a random cell line conditioned to be contaminated at generation n. The following example even proves that $\mathbb{P}^n(Z_{\mathbf{U}_n} \in .)$ and $\mathbb{P}(Z_n^* \in .)$ could be *a priori* very different.

Example 6. Suppose that generation n (fixed) contains 100 cells with 1 parasite (and no other contaminated cells) with probability 1/2 and it contains 1 cell with 100 parasites with probability 1/2 (and no other contaminated cells). Compare then

$$\mathbb{P}^n(Z_{\mathbf{U}_n} = 1) = 1/2, \qquad \mathbb{P}^n(Z_{\mathbf{U}_n} = 100) = 1/2;$$
$$\mathbb{P}(Z_n^* = 1) = 100/101, \qquad \mathbb{P}(Z_n^* = 100) = 1/101.$$

Actually the convergence of $(Z_n^*)_{n\in\mathbb{N}}$ leads to the result obtained by Kimmel [59] in the continuous analogue of this model. That is,

$$\frac{\mathbb{P}(Z_n = k)}{\mathbb{P}(Z_n > 0)} = \frac{\sum_{\mathbf{i}\in\mathbb{G}_n} \mathbb{E}(\mathbb{1}_{Z_\mathbf{i}=k})}{\sum_{\mathbf{i}\in\mathbb{G}_n} \mathbb{E}(\mathbb{1}_{Z_\mathbf{i}>0})} = \frac{\mathbb{E}(\#\{\mathbf{i}\in\mathbb{G}_n : Z_\mathbf{i} = k\})}{\mathbb{E}(\#\mathbb{G}_n^*)} \tag{8.20}$$

tends to $\mathbb{P}(\Upsilon = k)$ whereas we are here interested in the expectation of $F_k(n)$.

A sufficient condition to get the equality of the two distributions is that $\#\mathbb{G}_n^*$ is deterministic, which does not hold here. But in the case when $(m_0, m_1) \in D_3$, we shall prove that $\#\mathbb{G}_n^*$ is asymptotically proportional to $(m_0 + m_1)^n$ as $n \to \infty$ (forthcoming Proposition 8.6.3). This enables us to control $\mathbb{P}^n(Z_{\mathbf{U}_n} \in .)$ by the distribution of $\mathbb{P}(Z_n^* \in .)$. More precisely, it is sufficient to prove the separation of descendances of parasites (Proposition 8.6.4) and the control of filled-in cells (Lemma 8.6.5) using the results about the BPRE Z_n^*. These two results are the keys for Theorems 8.5.2, 8.5.5 and 8.5.7. Similarly, when $(m_0, m_1) \in D_5$, we already know that $\#\mathbb{G}_n^*$ is approximatively equivalent to 2^n. Then the fact that Z_n^* explodes as $n \to \infty$ (by Proposition 6.3.1) will ensure that the proportion of filled-in cells among contaminated cells tends to one (Theorem 8.5.1 below).

8.5.1 Case $(m_0, m_1) \in D_5$ $(m > 1)$

In that case, recall that conditionally on Ext^c, $\#\mathbb{G}_n^*$ is asymptotically proportional to 2^n (by Theorem 8.3.1). Moreover the contaminated cells become largely infected, as stated below.

Theorem 8.5.1. *Conditionally on* Ext^c*, for every* $k \in \mathbb{N}$*,* $F_k(n)$ *converges in probability to* 0 *as* $n \to \infty$*. i.e.*

$$\forall K, \epsilon > 0, \quad \mathbb{P}^*\Big(\frac{\#\{\mathbf{i}\in\mathbb{G}_n : Z_\mathbf{i} \geq K\}}{\#\mathbb{G}_n^*} \geq 1-\epsilon\Big) \overset{n\to\infty}{\longrightarrow} 1.$$

If $m_0 = m_1$, the number of parasites in a contaminated cell is of the same order as m_0^n. More precisely, for every $\epsilon > 0$,

$$\sup_{n\in\mathbb{N}} \Big\{\mathbb{P}^*\Big(\frac{\#\{\mathbf{i}\in\mathbb{G}_n^* : Z_\mathbf{i} \leq \alpha m_0^n\}}{\#\mathbb{G}_n^*} \geq \epsilon\Big)\Big\} \overset{\alpha\to 0}{\longrightarrow} 0.$$

Proof. In that case, use Theorem 8.3.1 and (8.19) to get that there exists a non negative random variable $\widetilde{L}$ such that

$$\#\mathbb{G}_n^* \geq 2^n \widetilde{L}, \qquad \mathbb{P}^*(\widetilde{L} = 0) = 0. \tag{8.21}$$

Let K, η and $\epsilon > 0$ and put $B_n(K,\eta) := \left\{ \frac{\#\{\mathbf{i}\in\mathbb{G}_n^* \,:\, Z_{\mathbf{i}}\leq K\}}{\#\mathbb{G}_n^*} \geq \eta \right\} \cap \mathrm{Ext}^c$, then

$$\sum_{\mathbf{i}\in\mathbb{G}_n^*} \mathbb{1}_{\{Z_{\mathbf{i}}\leq K\}} \geq \eta 2^n \widetilde{L}\mathbb{1}_{B_n(K,\eta)}$$

which gives, taking expectations,

$$\mathbb{E}(\widetilde{L}\mathbb{1}_{B_n(K,\eta)}) \leq \frac{\mathbb{E}(\sum_{\mathbf{i}\in\mathbb{G}_n^*} 2^{-n}\mathbb{1}_{\{Z_{\mathbf{i}}\leq K\}})}{\eta} = \frac{\mathbb{P}(0 < Z_n \leq K)}{\eta}.$$

Use then Proposition 6.3.1 in Chapter 6 and (8.21) to choose n large enough so that

$$\mathbb{P}(B_n(K,\eta)) \leq \epsilon,$$

which completes the proof of the theorem. In the case $m_0 = m_1 = m$, follow the proof above and use that Z_n/m^n converges to a positive limit on Ext^c (see [7]) to get the finer result given after the theorem. □

8.5.2 Case $(m_0, m_1) \in D_3$ $(m \leq 1)$ and simulations

We assume here $\mathbb{E}(Z^{(a)2}) < \infty$ and prove that $(F_k(n))_{k\in\mathbb{N}}$ converges to a deterministic limit. We prove the convergence thanks to the Cauchy criterion (using completeness of $l^1(\mathbb{N})$). The fact that the limit is deterministic is a consequence of the separation of the descendances of parasites and the law of large numbers. Once we know this limit is deterministic, we identify it with the Yaglom limit Υ (see Section 6.1 for proofs).

Theorem 8.5.2. *Conditionally on* Ext^c*, as* $n \to \infty$*,* $(F_k(n))_{k\in\mathbb{N}}$ *converges in probability in* $\mathbb{S}^1(\mathbb{N})$ *to* $(\mathbb{P}(\Upsilon = k))_{k\in\mathbb{N}}$.

Remark 13. We get here a realization of the Yaglom distribution Υ.

The limit just depends on the one-dimensional distributions of $(Z^{(0)}, Z^{(1)})$. More precisely, recall that the probability generating function G of Υ is characterized by (8.15).

This theorem still holds starting from k parasites. We get also easily a similar result in the case when a cell gives birth to N cells ($N \in \mathbb{N}$).

As an application, we can obtain numerically the Yaglom quasistationary distribution of any BGW. Let Z be the reproduction law of a BGW with mean $m < 1$ and choose N such that $Nm > 1$. Consider Kimmel's model where each cell divides into N daughter cells and $Z^{(0)} \stackrel{d}{=} Z^{(1)} \stackrel{d}{=} .. \stackrel{d}{=} Z^{(N)} \stackrel{d}{=} Z$. Computing then the asymptotic of the proportions of contaminated cells with k parasites gives the Yaglom quasistationary distribution associated to Z. If $\mathbb{P}(\mathrm{Ext}) \neq 0$, one can start from many parasites 'to avoid' extinction.

More generally, we can obtain similarly the Yaglom quasistationary distribution of any BPRE with finite number k of environments such that $\sum_1^k m_i^2 < \sum_1^k m_i$.

This theorem is in the same vein as Theorem 11 in [47]. But we can not follow the same approach as Guyon for the proof. Indeed we have to consider here the proportions among the contaminated cells in generation n whereas Guyon considers proportions among all cells in generation n. Unfortunately, the subtree of contaminated cells is itself random and induces long-range dependences between cells lines, so that Guyon's arguments do not hold here. Moreover Theorem 11 in [47] relies on an ergodicity hypothesis which cannot be circumvented.

Example 7. We give two examples when the limit can be calculated.

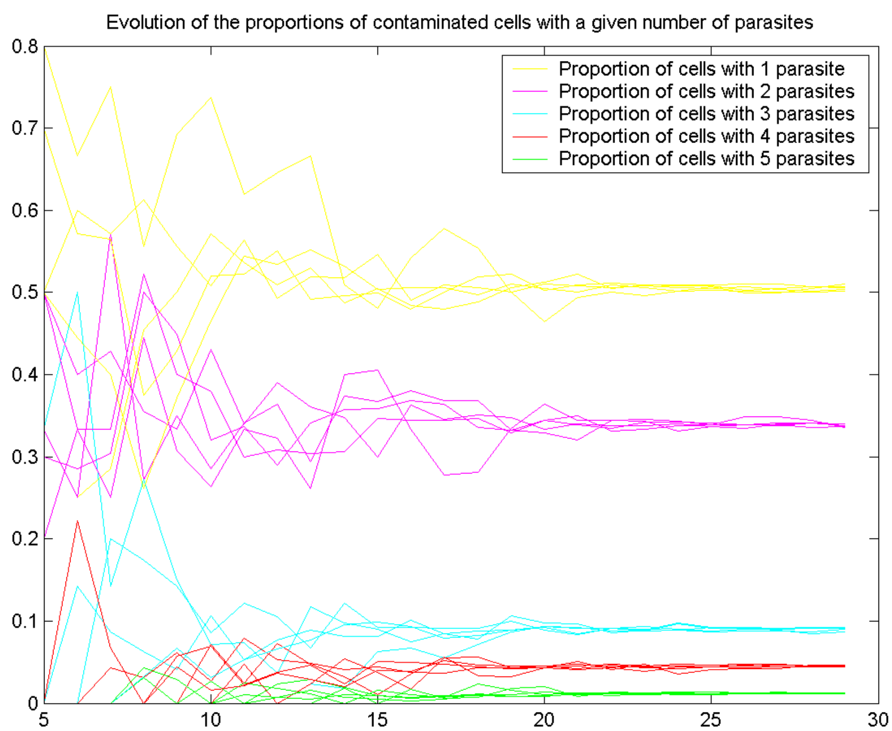
- $\star$ Trivial case : $\mathbb{P}(Z^{(0)} \in \{0,1\}, Z^{(1)} \in \{0,1\}) = 1$ leads to $\mathbb{P}(\Upsilon = 1) = 1$.
- $\star$ Symmetric linear fractional case : $p \in]0,1[, b \in]0,(1-p)^2[$ and

$$\mathbb{P}(Z^{(0)} = k) = \mathbb{P}(Z^{(1)} = k) = bp^{k-1} \text{ if } k \geq 1$$

and $\mathbb{P}(Z^{(0)} = 0) = \mathbb{P}(Z^{(1)} = 0) = (1-b-p)/(1-p)$. Then $m_0 = m_1 = b/(1-p)^2 < 1$ and letting s_0 be the root of $f_0(s) = s$ larger than 1,

$$\forall k \geq 1, \qquad \mathbb{P}(\Upsilon = k) = (s_0 - 1)/s_0^k.$$

Figure 6. Simulations for asymptotic behaviour of $F_k(n)$ (with $0 \leq Z^{(0)}, Z^{(1)} \leq 3$ a.s).

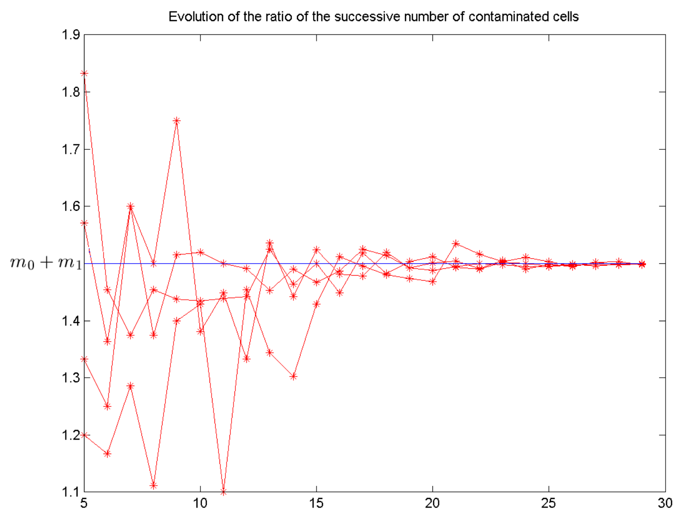


As asymptotically we know the number of parasites and the proportion of cells with k parasites, we get the number of contaminated cells (recall that W is given by (8.6)).

Corollary 8.5.3. *Conditionally on* Ext^c*, the following convergences hold in probability*

$$\frac{\#\mathbb{G}_n^*}{\mathcal{Z}_n} \xrightarrow{n\to\infty} \frac{1}{\mathbb{E}(\Upsilon)}, \qquad \frac{\#\mathbb{G}_n^*}{(m_0+m_1)^n} \xrightarrow{n\to\infty} \frac{W}{\mathbb{E}(\Upsilon)}.$$

Figure 7. Simulation for the asymptotic behavior of $\#\mathbb{G}_{n+1}^*/\#\mathbb{G}_n^*$ (with $0 \leq Z^{(0)}, Z^{(1)} \leq 3$ a.s.).



We can also consider the ancestors at generation n of the cells of $\partial\mathbb{T}^*$, which amounts to considering

$$F_k(n,p) = \frac{\#\{\mathbf{i} \in \mathbb{G}_{n+p}^* : Z_{\mathbf{i}|n} = k\}}{\#\mathbb{G}_{n+p}^*}$$

and let $p \to \infty$. Letting then $n \to \infty$ yields the biased Yaglom quasistationary distribution, thanks to the separation of descendances of parasites.

Corollary 8.5.4. *Conditionally on* Ext^c*, for every* $k \in \mathbb{N}$*,* $F_k(n,p)$ *converges in probability in* $\mathbb{S}^1(\mathbb{N})$ *as* p *tends to infinity. This limit converges in probability in* $\mathbb{S}^1(\mathbb{N})$ *as* $n \to \infty$*.*

$$\forall k \in \mathbb{N}, \qquad \lim_{n\to\infty}\lim_{p\to\infty} F_k(n,p) \overset{\mathbb{P}}{=} \frac{k\mathbb{P}(\Upsilon = k)}{\mathbb{E}(\Upsilon)}.$$

We get here an interpretation of the fact that the stationary distribution of the Q-process associated to the BPRE $(Z_n)_{n\in\mathbb{N}}$ is the size-biased Yaglom limit (see [1]).

8.5.3 Case $(m_0, m_1) \in D_2$

In that case, the parasites die out. So we condition by $\mathcal{Z}_n > 0$, we still assume $\mathbb{E}(Z^{(a)\ 2}) < \infty$ and we get a similar result.

Theorem 8.5.5. *As $n \to \infty$, $(F_k(n))_{k\in\mathbb{N}}$ conditioned by $\mathcal{Z}_n > 0$ converges in distribution on $\mathbb{S}^1(\mathbb{N})$ to $(\mathbb{P}(\Upsilon = k))_{k\in\mathbb{N}}$.*

The proof follows that of the previous theorem. Indeed (8.13) is still satisfied and we can use the same results on the BPRE $(Z_n)_{n\in\mathbb{N}}$. There are only two differences. First, we work under $\mathbb{P}^n$ instead of $\mathbb{P}^*$. Moreover $\mathcal{Z}_n$ satisfies now $\mathbb{P}(\mathcal{Z}_n > 0) \stackrel{n\to\infty}{\sim} 2/(\mathrm{Var}(Z^{(0)} + Z^{(1)})n)$ and $\mathcal{Z}_n/n$ conditioned to be non zero converges in distribution as $n \to \infty$ to an exponential variable $\mathcal{E}$ of parameter $2/(\hat{m}+1)$ (see section 2.1). As above, we can derive the following result.

Corollary 8.5.6. *As $n \to \infty$, $\#\mathbb{G}_n^*/n$ conditioned by $\#\mathbb{G}_n^* > 0$ converges in distribution to $\mathcal{E}/\mathbb{E}(\Upsilon)$.*

8.5.4 Case $(m_0, m_1) \in D_1$

In this case, the number of contaminated cells does not explode and the number of cells of type k at generation n conditioned by the survival of parasites in this generation converges weakly to a non deterministic limit (see Section 7 for proofs).

Theorem 8.5.7. *As $n \to \infty$, $(\#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} = k\})_{k\in\mathbb{N}}$ conditioned on $\mathcal{Z}_n > 0$ converges in distribution on $l^1(\mathbb{N})$ to a random sequence $(N_k)_{k\in\mathbb{N}}$ which satisfies $\mathbb{E}(\sum_{k\in\mathbb{N}} kN_k) < \infty$.*

As above, we get

Corollary 8.5.8. *$\#\mathbb{G}_n^*$ conditioned by $\#\mathbb{G}_n^* > 0$ converges in distribution to a positive finite random variable.*

Picking a cell uniformly on $\partial\mathbb{T}^*$ leads again to the size-biased distribution.

Corollary 8.5.9. *For every $n \in \mathbb{N}$, $(\#\{\mathbf{i} \in \mathbb{G}_{n+p}^* : Z_{\mathbf{i}|n} = k\})_{k\in\mathbb{N}}$ conditioned on $\mathcal{Z}_{n+p} > 0$ converges weakly in $l^1(\mathbb{N})$ to a random sequence as p tends to infinity. This limit converges weakly as $n \to \infty$.*

$$\forall k \in \mathbb{N}, \qquad \lim_{n\to\infty} \lim_{p\to\infty} \#\{\mathbf{i} \in \mathbb{G}_{n+p}^* : Z_{\mathbf{i}|n} = k\} \ |\mathcal{Z}_{n+p} > 0 = \frac{kN_k}{\sum_{k'\in\mathbb{N}} k' N_{k'}}.$$

8.5.5 Remaining domain : $(m_0, m_1) \in D_4$

In this domain, the asymptotic of the mean of the number of contaminated cells, that is $\mathbb{E}(\#\mathbb{G}_n^*) = 2^n\mathbb{P}(Z_n > 0)$, is different from the previous ones.
Recalling Section 2.2, this asymptotic depends on three subdomains, the interior of D_4 and the two connex components of its boundary. More precisely, it depends on $m_0m_1 = 1$ or $m_0m_1 < 1$ and $m_0\log(m_0) + m_1\log(m_1)$ is positive or zero.

If $(m_0, m_1) \in D_4$ and $m_0 < 1 < m_1$, using (8.16) and a coupling argument with Corollary 8.5.3, one can prove that

$$\sup_{n\in\mathbb{N}} \big\{\mathbb{P}\big(\frac{\#\mathbb{G}_n^*}{2^n\mathbb{P}(Z_n > 0)} \geq A,\ \frac{\#\mathbb{G}_n^*}{(m_0 + \widetilde{m_0})^n} \leq 1/A\big)\big\} \overset{A\to 0}{\longrightarrow} 0,$$

where $\widetilde{m_0} = \big(1 + \sqrt{1 + 4(m_0 - m_0^2)}\big)/2 > 1$. Thus $\#\mathbb{G}_n^*$ grows geometrically and one can naturally conjecture that $\#\mathbb{G}_n^*$ is asymptotically proportional to $\mathbb{E}(\#\mathbb{G}_n^*) = 2^n\mathbb{P}(Z_n > 0)$.

Moreover separation of descendances of parasites, control of filled-in cells and Corollary 8.5.4 do not hold in this case. Thus determining the limit behaviors here requires a different approach.

Finally, note that in the subdomain $m_0m_1 = 1$ (boundary of D_5), $(Z_n^*)_{n\in\mathbb{N}}$ explodes (see [**?**]) so the asymptotic proportion of contaminated cells which are arbitrarily largely contaminated should be equal to 1 as in Theorem 8.5.1.

8.6 Proofs in the case $(m_0, m_1) \in D_3$

We assume in this section that $\mathbb{E}(Z^{(a)2}) < \infty$ (i.e. $\widetilde{m} < \infty$) and we start with giving some technical results.

8.6.1 Preliminaries

First, note that for all $u, v \in l^1(\mathbb{N}^*)$, we have,

$$\|\frac{u}{\|u\|_1} - \frac{v}{\|v\|_1}\|_1 = \|\frac{u - v}{\|u\|_1} + \frac{v}{\|v\|_1}\frac{\|v\|_1 - \|u\|_1}{\|u\|_1}\|_1 \leq 2\frac{\|u - v\|_1}{\|u\|_1}. \tag{8.22}$$

Moreover by (8.6), there exist two random variables C and D a.s finite such that

$$\forall n \in \mathbb{N},\ C \leq \frac{\mathcal{Z}_n}{(2m)^n} \leq D \quad \text{a.s}, \qquad \mathbb{P}^*(C = 0) = \mathbb{P}^*(D = 0) = 0 \tag{8.23}$$

and as $\cap_{n\in\mathbb{N}}\{\mathcal{Z}_n > 0\} = \{\forall n \in \mathbb{N} : \mathcal{Z}_n > 0\}$, we have,

$$\sup_A\{\big|\mathbb{P}^n(A) - \mathbb{P}^*(A)\big|\} \overset{n\to\infty}{\longrightarrow} 0. \tag{8.24}$$

We focus now on the BPRE $(Z_n)_{n\in\mathbb{N}}$. First, by induction and convexity of f_a, we have for every $\mathbf{i} \in \mathbb{G}_n$, (see Section 2.2 for the Notation)

$$\mathbb{P}(Z_{\mathbf{i}} > 0) = 1 - f_{\mathbf{i}}(0) \leq m_{\mathbf{i}}. \tag{8.25}$$

Then identities (8.25) and (8.14) entail that there exists $M > 0$ such that

$$M \leq \frac{\mathbb{P}(Z_n > 0)}{m^n} \leq 1. \tag{8.26}$$

Moreover, by Corollary 2.3 in [1], we have

$$\lim_{K\to\infty} \sup_{n\in\mathbb{N}} \{\mathbb{E}(Z_n \mathbb{1}_{Z_n \geq K} \mid Z_n > 0)\} = 0. \tag{8.27}$$

Finally, Proposition 7.2.1 in Chapter 7 ensures that, if $(Z_n^{(1)})_{n\in\mathbb{N}}$ and $(Z_n^{(2)})_{n\in\mathbb{N}}$ are two independent BPRE distributed as $(Z_n)_{n\in\mathbb{N}}$, we have

$$\mathbb{P}(Z_n^{(1)} > 0,\ Z_n^{(2)} > 0) = o(\mathbb{P}(Z_n > 0)) = o(m^n) \quad (n \to \infty).$$

Then, we have

$$2^{-n} \sum_{\mathbf{i}\in\mathbb{G}_n} \mathbb{P}(Z_{\mathbf{i}} > 0)^2 = o(m^n) \quad (n \to \infty). \tag{8.28}$$

8.6.2 Estimation of $\#\mathbb{G}_n^*$

We prove here that the number of parasites which belong to filled-in cells is negligible compared to the total number of parasites (see also Lemma 8.6.5 for a result of the same kind). To prove this result, we use its counterpart for BPRE $(Z_n)_{n\in\mathbb{N}}$ conditioned to be non zero.

Lemma 8.6.1. *For every $\eta > 0$,*

$$\sup_{n\in\mathbb{N}} \left\{ \mathbb{P}^* \left(\frac{\sum_{\mathbf{i}\in\mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}} > K\}}}{\mathcal{Z}_n} \geq \eta \right) \right\} \xrightarrow{K\to\infty} 0.$$

Proof. Let $\eta > 0$ and write $A_n(K,\eta) := \left\{ \frac{\sum_{\mathbf{i}\in\mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}} > K\}}}{\mathcal{Z}_n} \geq \eta \right\} \cap \mathrm{Ext}^c$. Then

$$\mathbb{1}_{A_n(K,\eta)} \sum_{\mathbf{i}\in\mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}} > K\}} \geq \mathbb{1}_{A_n(K,\eta)} \mathcal{Z}_n \eta.$$

Using (8.23), we have,

$$\mathbb{1}_{A_n(K,\eta)} (2m)^{-n} \sum_{\mathbf{i}\in\mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}} > K\}} \geq \eta \mathbb{1}_{A_n(K,\eta)} C$$

so that taking expectations,

$$\begin{aligned} m^{-n}\mathbb{E}(2^{-n}\sum_{\mathbf{i}\in\mathbb{G}_n^*} Z_{\mathbf{i}}\mathbb{1}_{\{Z_{\mathbf{i}}>K\}}) &\geq \mathbb{E}(\mathbb{1}_{A_n(K,\eta)}C)\eta \\ m^{-n}\mathbb{E}(Z_n\mathbb{1}_{\{Z_n>K\}})/\eta &\geq \mathbb{E}(\mathbb{1}_{A_n(K,\eta)}C). \end{aligned}$$

Then, by (8.27), we have

$$\lim_{K\to\infty}\sup_{n\in\mathbb{N}}\{\mathbb{E}(\mathbb{1}_{A_n(K,\eta)}C)\} = 0.$$

Then observe that $\forall\alpha > 0$, $\inf_{\mathbb{P}^*(A)\geq\alpha}\{\mathbb{E}(C\mathbb{1}_A)\} > 0$. So $\exists K_0 \geq 0$ such that $\forall K \geq K_0$, $\forall n \in \mathbb{N}$,

$$\mathbb{P}^*(A_n(K,\eta)) < \alpha,$$

which completes the proof. $\square$

First, for any $\epsilon > 0$, choose K using the previous lemma such that

$$\mathbb{P}^*\left(\frac{\sum_{\mathbf{i}\in\mathbb{G}_n^*} Z_{\mathbf{i}}\mathbb{1}_{\{Z_{\mathbf{i}}\leq K\}}}{\mathcal{Z}_n} \geq 1/2\right) = 1 - \mathbb{P}^*\left(\frac{\sum_{\mathbf{i}\in\mathbb{G}_n^*} Z_{\mathbf{i}}\mathbb{1}_{\{Z_{\mathbf{i}}>K\}}}{\mathcal{Z}_n} < 1/2\right) \geq 1 - \epsilon/2.$$

Adding that conditionally on Ext^c, $\mathcal{Z}_n \stackrel{n\to\infty}{\longrightarrow} \infty$ a.s, gives the following result.

Proposition 8.6.2. *Let $\epsilon > 0$, there exists $K \in \mathbb{N}$ such that $\forall N \in \mathbb{N}$, $\exists n_0 \in \mathbb{N}$ such that $\forall n \geq n_0$,*

$$\mathbb{P}^*(\sum_{\mathbf{i}\in\mathbb{G}_n^*} Z_{\mathbf{i}}\mathbb{1}_{\{Z_{\mathbf{i}}\leq K\}} \geq N) \geq 1 - \epsilon.$$

Second, we derive an estimation of $\#\mathbb{G}_n^*$. By Lemma 8.6.1, the cells are not very contaminated so the number of contaminated cells is asymptotically proportional to the number of parasites, which is a a Bienaymé Galton Watson process.

Proposition 8.6.3. *For every $\epsilon > 0$, there exist $A, B > 0$ such that for every $n \in \mathbb{N}$,*

$$\mathbb{P}^*\left(\frac{\#\mathbb{G}_n^*}{(2m)^n} \in [A,B]\right) \geq 1 - \epsilon.$$

Proof. First use (8.23) to get

$$\frac{\#\mathbb{G}_n^*}{(2m)^n} \leq \frac{\mathcal{Z}_n}{(2m)^n} \leq D.$$

Moreover using again (8.23), we have

$$\frac{\#\mathbb{G}_n^*}{(2m)^n} \geq \frac{\sum_{\mathbf{i}\in\mathbb{G}_n^*} Z_{\mathbf{i}}\mathbb{1}_{\{Z_{\mathbf{i}}\leq K\}}}{K(2m)^n} \geq \frac{C}{K}\frac{\sum_{\mathbf{i}\in\mathbb{G}_n^*} Z_{\mathbf{i}}\mathbb{1}_{\{Z_{\mathbf{i}}\leq K\}}}{\mathcal{Z}_n}$$

and Lemma 8.6.1 gives the result. $\square$

8.6.3 Separation of the descendances of parasites

Start with two parasites and consider the BPRE $(Z_n)_{n\in\mathbb{N}}$. Even when conditioning on the survival of their descendance, the descendance of one of them dies out. This ensures that two distinct parasites in generation n do not have descendants which belong to the same cell in generation $n+q$ if q is large enough. More precisely, we define $N_n(\mathbf{i})$ as the number of parasites of cell $\mathbf{i}|n$ whose descendance is still alive in cell $\mathbf{i}$ and we prove the following result.

Proposition 8.6.4. *$\forall K \in \mathbb{N},\ \forall\ \epsilon, \eta > 0,\ \exists q \in \mathbb{N}$ such that $\forall n \in \mathbb{N}$, we have*

$$\mathbb{P}^*\bigg(\frac{\#\{\mathbf{i} \in \mathbb{G}^*_{n+q} : Z_{\mathbf{i}|n} \leq K,\ N_n(\mathbf{i}) \geq 2\}}{\#\mathbb{G}^*_{n+q}} \geq \eta\bigg) \leq \epsilon.$$

Proof. Let $K \in \mathbb{N},\ \eta > 0$ and consider for $A > 0$,

$$E^q_n(\eta) = \bigg\{\frac{\#\{\mathbf{i} \in \mathbb{G}^*_{n+q} : Z_{\mathbf{i}|n} \leq K,\ N_n(\mathbf{i}) \geq 2\}}{\#\mathbb{G}^*_{n+q}} \geq \eta\bigg\} \cap \bigg\{\frac{\#\mathbb{G}^*_{n+q}}{(2m)^{n+q}} \geq A\bigg\}.$$

Then

$$\mathbb{1}_{E^q_n(\eta)}\#\{\mathbf{i} \in \mathbb{G}^*_{n+q} : Z_{\mathbf{i}|n} \leq K,\ N_n(\mathbf{i}) \geq 2\} \geq \mathbb{1}_{E^q_n(\eta)}\eta A(2m)^{n+q}$$

so that taking expectations,

$$\begin{aligned}
\mathbb{P}(E^q_n(\eta)) &\leq \frac{2^{-(n+q)}\mathbb{E}(\sum_{\mathbf{i}\in\mathbb{G}_{n+q}} \mathbb{1}_{\{Z_{\mathbf{i}|n}\leq K,\ N_n(\mathbf{i})\geq 2\}})}{\eta A m^{n+q}} \\
&\leq \frac{2^{-n}\sum_{\mathbf{j}\in\mathbb{G}_n}\mathbb{P}(0 < Z_{\mathbf{j}} \leq K)2^{-q}\sum_{\mathbf{i}\in\mathbb{G}_q}\mathbb{P}_K(N_0(\mathbf{i}) \geq 2)}{\eta A m^{n+q}} \\
&\leq \frac{\mathbb{P}(Z_n > 0)2^{-q}\sum_{\mathbf{i}\in\mathbb{G}_q}\mathbb{P}_K(N_0(\mathbf{i}) \geq 2)}{\eta A m^{n+q}}
\end{aligned}$$

As we have $\binom{K}{2}$ ways to choose two parasites among K and they both survive along $\mathbf{i}$ with probability $\mathbb{P}(Z_{\mathbf{i}} > 0)^2$, we have

$$\mathbb{P}_K(N_0(\mathbf{i}) \geq 2) \leq \binom{K}{2}\mathbb{P}(Z_{\mathbf{i}} > 0)^2.$$

Then

$$\mathbb{P}(E^q_n(\eta)) \leq \frac{\binom{K}{2}2^{-q}\sum_{\mathbf{i}\in\mathbb{G}_q}\mathbb{P}(Z_{\mathbf{i}} > 0)^2}{\eta A m^q}.$$

Conclude choosing A in agreement with Proposition 8.6.3 and q with (8.28). $\square$

8.6.4 Control of filled-in cells

Here we prove that filled-in cells have asymptotically no impact on the proportions of cells with a given number of parasites.

Lemma 8.6.5. *$\forall\ \epsilon, \eta > 0,\ \exists K \in \mathbb{N}$ such that $\forall\ n, q \in \mathbb{N}$, we have*

$$\mathbb{P}^*\Big(\frac{\#\{\mathbf{i} \in \mathbb{G}^*_{n+q} : Z_{\mathbf{i}|n} > K\}}{\#\mathbb{G}^*_{n+q}} \geq \eta\Big) \leq \epsilon.$$

Proof. Let $\eta > 0, A > 0$ and consider

$$F^q_n(\eta) = \left\{\frac{\#\{\mathbf{i} \in \mathbb{G}^*_{n+q} : Z_{\mathbf{i}|n} > K\}}{\#\mathbb{G}^*_{n+q}} \geq \eta\right\} \cap \left\{\frac{\#\mathbb{G}^*_{n+q}}{(2m)^{n+q}} \geq A\right\}.$$

then

$$\mathbb{1}_{F^q_n(\eta)}\#\{\mathbf{i} \in \mathbb{G}^*_{n+q} : Z_{\mathbf{i}|n} > K\} \geq \mathbb{1}_{F^q_n(\eta)}\eta A(2m)^{n+q}.$$

Taking expectations leads to

$$\begin{aligned}
\mathbb{P}(F^q_n(\eta)) &\leq \frac{2^{-(n+q)}\mathbb{E}(\sum_{\mathbf{i}\in\mathbb{G}_{n+q}} \mathbb{1}_{\{Z_{\mathbf{i}|n}>K, Z_{\mathbf{i}}>0\}})}{\eta A m^{n+q}} \\
&\leq \frac{2^{-(n+q)}\sum_{\mathbf{i}\in\mathbb{G}_{n+q}} \mathbb{P}(Z_{\mathbf{i}|n} > K, Z_{\mathbf{i}} > 0)}{\eta A m^{n+q}} \\
&\leq \frac{\sum_{k>K} 2^{-n}\sum_{\mathbf{j}\in\mathbb{G}_n} \mathbb{P}(Z_{\mathbf{j}} = k)2^{-q}\sum_{\mathbf{i}\in\mathbb{G}_q} \mathbb{P}_k(Z_{\mathbf{i}} > 0)}{\eta A m^{n+q}}
\end{aligned}$$

Moreover $\mathbb{P}_k(Z_{\mathbf{i}} > 0) = 1 - (1 - \mathbb{P}(Z_{\mathbf{i}} > 0))^k \leq k\mathbb{P}(Z_{\mathbf{i}} > 0)$ and we have

$$\begin{aligned}
\mathbb{P}(F^q_n(\eta)) &\leq \frac{\sum_{k>K} 2^{-n}\sum_{\mathbf{j}\in\mathbb{G}_n} k\mathbb{P}(Z_{\mathbf{j}} = k)\mathbb{P}(Z_q > 0)}{\eta A m^{n+q}} \\
&\leq \frac{\mathbb{E}(Z_n\mathbb{1}_{\{Z_n>K\}})}{\eta A m^n} \qquad \text{using (8.26).}
\end{aligned}$$

By (8.28), we get

$$\lim_{K\to\infty}\sup_{n\in\mathbb{N}}\{\mathbb{P}(F^q_n(\eta))\} = 0.$$

Complete the proof choosing A in agreement with Proposition 8.6.3. □

Proof of Theorem 8.5.2

Consider the contaminated cells in generation $n+q$. Their ancestors in generation n are cells which are not very contaminated (by Lemma 8.6.5). Then taking q large, the parasites of a contaminated cell in generation $n+q$ come from a same parasite in generation n (separation of the descendances of parasites, Proposition

8.6.4). Thus at generation $n+q$, everything occurs as if all parasites from generation n belonged to different cells. As the number of parasites at generation n tends to infinity ($n \to \infty,\ m_0 + m_1 > 1$), we have a law of large numbers phenomenon and get a deterministic limit.

STEP 1 : We prove that for all $\epsilon, \eta > 0$, there exist $n_0 \in \mathbb{N}$ and $\vec{f} \in \mathbb{S}^1(\mathbb{N})$ such that for every $n \geq n_0$,

$$\mathbb{P}^*(\|(F_k(n))_{k\in\mathbb{N}} - \vec{f}\|_1 \geq \eta) \leq \epsilon.$$

• For every $k \in \mathbb{N}^*$ and every parasite $\mathbf{p}$ in generation n, we denote by $Y_k^q(\mathbf{p})$ the number of cells in generation $n+q$ which contain at least k parasites, exactly k of which have $\mathbf{p}$ as an ancestor. By convention, $Y_0^q(\mathbf{p}) = 0$. That is, writing for $\mathbf{p}$ parasite, $\mathbf{p} \hookrightarrow \mathbf{i}$ when $\mathbf{p}$ belongs to the cell $\mathbf{i}$ and $\mathbf{p}|n$ its ancestor (parasite) in generation n,

$$Y_k^q(\mathbf{p}) = \sum_{\mathbf{i}\in\mathbb{G}_{n+q}} \mathbb{1}_{\#\{\mathbf{r}\,:\,\mathbf{r}\hookrightarrow\mathbf{i},\ \mathbf{r}|n=\mathbf{p}\}=k}, \qquad k \in \mathbb{N}^*.$$

By the branching property, $\big(Y_k^q(\mathbf{p})\big)_{k\in\mathbb{N}}$ $(\mathbf{p} \in \mathcal{P}(n))$ are iid and we denote by $(Y_k^q)_{k\in\mathbb{N}}$ a random variable with this common distribution. Denoting by $\mathcal{P}_K(n)$ the set of parasites in generation n which belong to a cell containing at most K parasites, we have

$$\begin{aligned} &\sum_{k\in\mathbb{N}^*} \Big|\#\{\mathbf{i} \in \mathbb{G}^*_{n+q} : Z_\mathbf{i} = k\} - \sum_{\mathbf{p}\in\mathcal{P}_K(n)} Y_k^q(\mathbf{p})\Big| \\ \leq\ & (K+1)\#\{\mathbf{i} \in \mathbb{G}^*_{n+q} : Z_{\mathbf{i}|n} \leq K,\ N_n(\mathbf{i}) \geq 2\} + \#\{\mathbf{i} \in \mathbb{G}^*_{n+q} : Z_{\mathbf{i}|n} > K\} \end{aligned} \tag{8.29}$$

Indeed, the left hand side of (8.29) is less than

$$\sum_{k\in\mathbb{N}^*} \Big|\#\{\mathbf{i} \in \mathbb{G}^*_{n+q} : Z_\mathbf{i} = k, Z_{\mathbf{i}|n} \leq K\} - \sum_{\mathbf{p}\in\mathcal{P}_K(n)} Y_k^q(\mathbf{p})\Big| + \#\{\mathbf{i} \in \mathbb{G}^*_{n+q} : Z_{\mathbf{i}|n} > K\}$$

And recalling that $N_n(\mathbf{i})$ is the number of parasites of cell $\mathbf{i}|n$ whose descendance is still alive in cell $\mathbf{i}$, we get the following equalities

$$\begin{aligned} \sum_{\mathbf{p}\in\mathcal{P}_K(n)} Y_k^q(\mathbf{p}) &= \sum_{\mathbf{i}\in\mathbb{G}_{n+q}} \sum_{\mathbf{p}\in\mathcal{P}_K(n)} \mathbb{1}_{\#\{\mathbf{r}\,:\,\mathbf{r}\hookrightarrow\mathbf{i},\ \mathbf{r}|n=\mathbf{p}\}=k} \\ \mathbb{1}_{Z_\mathbf{i}=k,\ Z_{\mathbf{i}|n}\leq K,\ N_n(\mathbf{i})=1} &= \mathbb{1}_{N_n(\mathbf{i})=1} \sum_{\mathbf{p}\in\mathcal{P}_K(n)} \mathbb{1}_{\#\{\mathbf{r}\,:\,\mathbf{r}\hookrightarrow\mathbf{i},\ \mathbf{r}|n=\mathbf{p}\}=k} \end{aligned}$$

which ensure

$$\sum_{k\in\mathbb{N}^*} \Big|\#\{\mathbf{i} \in \mathbb{G}^*_{n+q} : Z_\mathbf{i} = k,\ Z_{\mathbf{i}|n} \leq K\} - \sum_{\mathbf{p}\in\mathcal{P}_K(n)} Y_k^q(\mathbf{p})\Big|$$

$$\begin{aligned}
&\leq \sum_{k\in\mathbb{N}^*} \sum_{\mathbf{i}\in\mathbb{G}_{n+q}, N_n(\mathbf{i})\geq 2} \Big| \mathbb{1}_{Z_{\mathbf{i}}=k,\ Z_{\mathbf{i}|n}\leq K} - \sum_{\mathbf{p}\in\mathcal{P}_K(n)} \mathbb{1}_{\#\{\mathbf{r}\ :\ \mathbf{r}\hookrightarrow\mathbf{i},\ \mathbf{r}|n=\mathbf{p}\}=k} \Big| \\
&\leq \#\{\mathbf{i}\in\mathbb{G}^*_{n+q} : Z_{\mathbf{i}|n}\leq K,\ N_n(\mathbf{i})\geq 2\} + \sum_{\substack{\mathbf{i}\in\mathbb{G}_{n+q}, N_n(\mathbf{i})\geq 2\\ \mathbf{p}\in\mathcal{P}_K(n)}} \mathbb{1}_{\#\{\mathbf{r}\ :\ \mathbf{r}\hookrightarrow\mathbf{i},\ \mathbf{r}|n=\mathbf{p}\}>0} \\
&\leq \#\{\mathbf{i}\in\mathbb{G}^*_{n+q} : Z_{\mathbf{i}|n}\leq K,\ N_n(\mathbf{i})\geq 2\} + \sum_{\mathbf{i}\in\mathbb{G}_{n+q}, N_n(\mathbf{i})\geq 2} K\mathbb{1}_{Z_{\mathbf{i}|n}\leq K} \\
&= (K+1)\#\{\mathbf{i}\in\mathbb{G}^*_{n+q} : Z_{\mathbf{i}|n}\leq K,\ N_n(\mathbf{i})\geq 2\}
\end{aligned}$$

We shall now prove that the quantities on the right hand side of (8.29) are small when n and q are large enough and that $\sum_{\mathbf{p}\in\mathcal{P}_K(n)} Y_k^q(\mathbf{p})$ follow a law of lage number. To that purpose, let $\epsilon,\eta>0$ and for all $K,k,n,q\geq 0$ define

$$G_k^K(n,q) := \frac{\sum_{\mathbf{p}\in\mathcal{P}_K(n)} Y_k^q(\mathbf{p})}{\sum_{k\in\mathbb{N}}\sum_{\mathbf{p}\in\mathcal{P}_K(n)} Y_k^q(\mathbf{p})}.$$

• First, by Proposition 8.6.2 and (8.24), $\exists K_1\in\mathbb{N}$ such that $\forall N\in\mathbb{N}$, $\exists n_1\in\mathbb{N}$ such that $\forall K\geq K_1$, $\forall n\geq n_1$,

$$\mathbb{P}^n(|\mathcal{P}_K(n)|\geq N)\geq 1-\epsilon. \tag{8.30}$$

Moreover by Lemma 8.6.5, $\exists\ K_2\geq K_1$ such that $\forall n,q\in\mathbb{N}$,

$$\mathbb{P}^*\Big(\frac{\#\{\mathbf{i}\in\mathbb{G}^*_{n+q} : Z_{\mathbf{i}|n}>K_2\}}{\#\mathbb{G}^*_{n+q}}\geq\eta\Big)\leq\epsilon. \tag{8.31}$$

And by Proposition 8.6.4, $\exists q_0\in\mathbb{N}$ such that $\forall n\in\mathbb{N}$

$$\mathbb{P}^*\big(\frac{\#\{\mathbf{i}\in\mathbb{G}^*_{n+q_0} : Z_{\mathbf{i}|n}\leq K_2,\ N_n(\mathbf{i})\geq 2\}}{\#\mathbb{G}^*_{n+q_0}}\geq\eta/(K_2+1)\big)\leq\epsilon. \tag{8.32}$$

Use then (8.29), (8.31) and (8.32) to get

$$\mathbb{P}^*\Big(\frac{\sum_{k\in\mathbb{N}^*}\big|\#\{\mathbf{i}\in\mathbb{G}_{n+q_0} : Z_{\mathbf{i}}=k\}-\sum_{\mathbf{p}\in\mathcal{P}_{K_2}(n)} Y_k^{q_0}(\mathbf{p}))\big|}{\#\mathbb{G}^*_{n+q_0}}\geq 2\eta\Big)\leq 2\epsilon.$$

Then by (8.22), for every $n\in\mathbb{N}$, we have

$$\mathbb{P}^*\big(\|(F_k(n+q_0))_{k\in\mathbb{N}}-(G_k^{K_2}(n,q_0))_{k\in\mathbb{N}}\|_1\geq 4\eta\big)\leq 2\epsilon. \tag{8.33}$$

• Second, conditionally on $\mathcal{Z}_n>0$, $Y_k^{q_0}(\mathbf{p})$ ($\mathbf{p}\in\mathcal{P}_{K_2}(n)$) are iid. Then the law of large numbers (LLN) ensures that $\forall k\in\mathbb{N}$, as n and so $\mathcal{P}_{K_2}(n)$ becomes large :

$$G_k^{K_2}(n,q_0)\longrightarrow f_k(q_0) \qquad \text{where} \quad f_k(q_0) := \frac{\mathbb{E}(Y_k^{q_0})}{\sum_{k'\in\mathbb{N}}\mathbb{E}(Y_{k'}^{q_0})}.$$

To see that, divide the numerator and denominator of $G_k^{K_2}(n, q_0)$ by $\#\mathcal{P}_{K_2}(n)$. More precisely, by the LLN, there exists $N > 0$ such that for all $n \in \mathbb{N}$,

$$\mathbb{P}^n\big(\|(G_k^{K_2}(n, q_0))_{k\in\mathbb{N}^*} - \vec{f}(q_0)\|_1 \geq \eta,\ \mathcal{P}_{K_2}(n) \geq N\big) \leq \epsilon.$$

So using (8.30), there exists $n_1 \in \mathbb{N}$ such that for every $\forall\ n \geq n_1$,

$$\mathbb{P}^n\big(\|(G_k^{K_2}(n, q_0))_{k\in\mathbb{N}^*} - \vec{f}(q_0)\|_1 \geq \eta\big) \leq 2\epsilon.$$

Finally by (8.24), there exists $n_2 \geq n_1$ such that for every $n \geq n_2$,

$$\mathbb{P}^*\big(\|(G_k^{K_2}(n, q_0))_{k\in\mathbb{N}^*} - \vec{f}(q_0)\|_1 \geq \eta\big) \leq 3\epsilon. \tag{8.34}$$

As a conclusion, using (8.33) and (8.34), we have proved that for all $\epsilon, \eta > 0$, anf for every $n \geq n_2 + q_0$,

$$\mathbb{P}^*(\|(F_k(n))_{k\in\mathbb{N}^*} - \vec{f}(q_0)\|_1 \geq 5\eta) \leq 3\epsilon.$$

STEP 2 : Existence of the limit.

For every $l \in \mathbb{N}$, there exist $n_0(l) \in \mathbb{N}$ and $\vec{f}(l) \in \mathbb{S}^1(\mathbb{N})$ such that for every $n \geq n_0(l)$

$$\mathbb{P}(\|F(n) - \vec{f}(l)\|_1 \geq 1/2^{l+1}) \leq 1/2^l.$$

Then for all l, l' such that $2 \leq l \leq l'$: $\|\vec{f}(l') - \vec{f}(l)\|_1 \leq 1/2^l$ and completeness of $l^1(\mathbb{N})$ ensures that $(\vec{f}(l))_{l\in\mathbb{N}}$ converges in $\mathbb{S}^1(\mathbb{N})$ to a limit $\vec{f}$. Moreover $\|\vec{f}(l) - \vec{f}\|_1 \leq 1/2^l$ so for every $n \geq n_0(l)$,

$$\mathbb{P}(\|F(n) - \vec{f}\|_1 \geq 1/2^l) \leq 1/2^l$$

which ensures the convergence in probability of $(F_k(n))_{n\in\mathbb{N}}$ to $\vec{f}$ as $n \to \infty$.

STEP 3 : Characterization of the limit as $f_k = \mathbb{P}(\Upsilon = k)$.

By Proposition 8.2.2, we have

$$\forall k \in \mathbb{N} \qquad \mathbb{P}(Z_n = k \ \mid\ Z_n \neq 0) \overset{n\to\infty}{\longrightarrow} \mathbb{P}(\Upsilon = k) \tag{8.35}$$

Moreover for every $k \in \mathbb{N}^*$, using (8.20),

$$\mathbb{P}(Z_n = k \ \mid\ Z_n \neq 0) = \frac{\mathbb{E}(\#\{\mathbf{i} \in \mathbb{G}_n : Z_{\mathbf{i}} = k\})}{\mathbb{E}(\#\mathbb{G}_n^*)} = \frac{\mathbb{E}(F_k(n)\#\mathbb{G}_n^*)}{\mathbb{E}(\#\mathbb{G}_n^*)}.$$

As $F_k(n)$ converges in probability to a deterministic limit f_k, we get

$$\forall k \in \mathbb{N}, \qquad \mathbb{P}(Z_n = k \ \mid\ Z_n \neq 0) \overset{n\to\infty}{\longrightarrow} f_k. \tag{8.36}$$

Indeed, by Proposition 8.6.3, there exists $A > 0$ such that

$$\frac{\mathbb{E}(\#\mathbb{G}_n^*)}{(2m)^n} \geq A.$$

Then for every $\eta > 0$, using $\mid F_k(n) - f_k \mid \leq 1$, we have

$$\begin{aligned}
\left|\frac{\mathbb{E}(F_k(n)\#\mathbb{G}_n^*)}{\mathbb{E}(\#\mathbb{G}_n^*)} - f_k\right| &\leq \frac{\mathbb{E}(\#\mathbb{G}_n^* \mid F_k(n) - f_k \mid \mathbb{1}_{\{|F_k(n)-f_k|<\eta\}})}{\mathbb{E}(\#\mathbb{G}_n^*)} \\
&\quad + \frac{\mathbb{E}(\#\mathbb{G}_n^* \mathbb{1}_{\{|F_k(n)-f_k|\geq\eta\}})}{\mathbb{E}(\#\mathbb{G}_n^*)} \\
&\leq \eta + \frac{\mathbb{E}(\mathcal{Z}_n \mathbb{1}_{\{|F_k(n)-f_k|\geq\eta\}})}{A(2m)^n}
\end{aligned}$$

By (8.11), $\mathcal{Z}_n/(2m)^n$ is bounded in $L^2$and it is uniformly integrable. Then, thanks to the previous steps, the second term in the last displayed equation vanishes as n grows and we get (8.36). Putting (8.35) and (8.36) together proves that $f_k = \mathbb{P}(\Upsilon = k)$.

<u>Proof of corollaries</u>

Proof of Corollary 8.5.3. Recall that $\mathbb{E}(\Upsilon) < \infty$ (Proposition 8.2.2) and note also that for every $K \in \mathbb{N}^*$,

$$\#\mathbb{G}_n^* = \frac{\sum_{\mathbf{i}\in\mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}}\leq K\}}}{\sum_{k=1}^K kF_k(n)}.$$

Then using $\sum_{\mathbf{i}\in\mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}}\leq K\}} \leq \mathcal{Z}_n$ gives

$$\begin{aligned}
\left|\frac{\#\mathbb{G}_n^*}{\mathcal{Z}_n} - \frac{1}{\mathbb{E}(\Upsilon)}\right| &= \left|\frac{1}{\sum_{k=1}^K kF_k(n)} \frac{\sum_{\mathbf{i}\in\mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}}\leq K\}}}{\mathcal{Z}_n} - \frac{1}{\mathbb{E}(\Upsilon)}\right| \\
&\leq \left|\frac{1}{\sum_{k=1}^K kF_k(n)} - \frac{1}{\mathbb{E}(\Upsilon)}\right| + \frac{1}{\mathbb{E}(\Upsilon)}\left|\frac{\sum_{\mathbf{i}\in\mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}}\leq K\}}}{\mathcal{Z}_n} - 1\right|
\end{aligned}$$

Let $\eta, \epsilon > 0$. We use Lemma 8.6.1 to choose $K \in \mathbb{N}^*$ such that

$$\forall n \in \mathbb{N},\ \mathbb{P}^*\left(\frac{\sum_{\mathbf{i}\in\mathbb{G}_n^*} Z_{\mathbf{i}} \mathbb{1}_{\{Z_{\mathbf{i}}\leq K\}}}{\mathcal{Z}_n} \geq 1 - \eta\right) \geq 1 - \epsilon \quad ; \quad \left|\frac{1}{\mathbb{E}(\Upsilon \mathbb{1}_{\Upsilon\leq K})} - \frac{1}{\mathbb{E}(\Upsilon)}\right| \leq \eta.$$

Choose $n_0 \in \mathbb{N}$ using Theorem 8.5.2 so that for every $n \geq n_0$,

$$\mathbb{P}^*\left(\left|\frac{1}{\sum_{k=1}^K kF_k(n)} - \frac{1}{\mathbb{E}(\Upsilon \mathbb{1}_{\Upsilon\leq K})}\right| \leq \eta\right) \geq 1 - \epsilon.$$

Then for every $n \geq n_0$,

$$\mathbb{P}^*\left(\left|\frac{\#\mathbb{G}_n^*}{\mathcal{Z}_n} - \frac{1}{\mathbb{E}(\Upsilon)}\right| \geq 2\eta + \frac{1}{\mathbb{E}(\Upsilon)}\eta\right) \leq 2\epsilon,$$

which proves the convergence in probability of $\#\mathbb{G}_n^*/\mathcal{Z}_n$ to $1/\mathbb{E}(\Upsilon)$. The second convergence follows from (8.6). $\square$

Proof of Corollary 8.5.4. We write for $n, p, k \in \mathbb{N}$,

$$\frac{\#\{\mathbf{i} \in \mathbb{G}^*_{n+p} : Z_{\mathbf{i}|n} = k\}}{\#\mathbb{G}^*_{n+p}} = \frac{(2m)^p}{\#\mathbb{G}^*_{n+p}} \sum_{\mathbf{j} \in \mathbb{G}^*_n : Z_{\mathbf{j}} = k} \frac{\#\{\mathbf{i} \in \mathbb{G}^*_{n+p} : \mathbf{i}|n = \mathbf{j}\}}{(2m)^p}.$$

Conditionally on $Z_{\mathbf{j}} = k$, by Corollary 8.5.3 and separation of descendances of parasites, we have the following convergence in probability

$$\frac{\#\{\mathbf{i} \in \mathbb{G}^*_{n+p} : \mathbf{i}|n = \mathbf{j}\}}{(2m)^p} \xrightarrow{p \to \infty} W_k(\mathbf{j}),$$

where $W_k(\mathbf{j})$ is the sum of k iid variables distributed as $W/\mathbb{E}(\Upsilon)$. Then, using also (8.6),

$$\mathbb{E}(W_k(\mathbf{j})) = \frac{k\mathbb{E}(W)}{\mathbb{E}(\Upsilon)} = \frac{k}{\mathbb{E}(\Upsilon)}. \tag{8.37}$$

Using again Corollary 8.5.3, we get the first limit of the Corollary

$$\lim_{p\to\infty} \frac{\#\{\mathbf{i} \in \mathbb{G}^*_{n+p} : Z_{\mathbf{i}|n} = k\}}{\#\mathbb{G}^*_{n+p}} \overset{\mathbb{P}}{=} \frac{\mathbb{E}(\Upsilon)}{W} \frac{\sum_{\mathbf{j} \in \mathbb{G}^*_n : Z_{\mathbf{j}} = k} W_k(\mathbf{j})}{(2m)^n}.$$

Moreover, Theorem 8.5.2 ensures that

$$\frac{\#\{\mathbf{j} \in \mathbb{G}^*_n : Z_{\mathbf{j}} = k\}}{(2m)^n} = F_k(n) \frac{\mathcal{Z}_n}{(2m)^n} \xrightarrow{n\to\infty} \frac{W}{\mathbb{E}(\Upsilon)} f_k.$$

And conditionally on $\#\mathbb{G}^*_n > 0$, $W_k(\mathbf{j})$ ($\mathbf{j} \in \mathbb{G}^*_n$) is idd by the branching property and $\#\mathbb{G}^*_n$ tends to infinity. So the law of large numbers and (8.37) ensure that

$$\begin{aligned}
&\lim_{n\to\infty} \frac{\mathbb{E}(\Upsilon)}{W} \frac{\sum_{\mathbf{j} \in \mathbb{G}^*_n : Z_{\mathbf{j}} = k} W_k(\mathbf{j})}{(2m)^n} \\
= \; &\lim_{n\to\infty} \frac{\mathbb{E}(\Upsilon)}{W} \frac{\#\{\mathbf{j} \in \mathbb{G}^*_n : Z_{\mathbf{j}} = k\}}{(2m)^n} \frac{\sum_{\mathbf{j} \in \mathbb{G}^*_n : Z_{\mathbf{j}} = k} W_k(\mathbf{j})}{\#\{\mathbf{j} \in \mathbb{G}^*_n : Z_{\mathbf{j}} = k\}} \overset{\mathbb{P}^*}{=} \frac{k f_k}{\mathbb{E}(\Upsilon)},
\end{aligned}$$

which ends the proof. □

8.7 Proofs in the case $(m_0, m_1) \in D_1$

We still assume $\mathbb{E}\big(Z^{(a)2}\big) < \infty$, the proof is in the same vein as the proof in the previous section and use the separation of the descendances of the parasites. The main difference is that $\mathcal{Z}_n$ does not explode so the limit is not deterministic and the convergence holds in distribution.

Lemma 8.7.1. *For every $K > 0$, there exists $q_0 \in \mathbb{N}$ such that for all $q \geq q_0$ and $n \in \mathbb{N}$,*

$$\mathbb{P}^{n+q}\big(\{\mathbf{i} \in \mathbb{G}^*_{n+q} : \ N_n(\mathbf{i}) \geq 2\} \neq \varnothing, \ \mathcal{Z}_n \leq K\big) \leq \epsilon.$$

Proof. Denotng by E^q_n the event

$$\{\{\mathbf{i} \in \mathbb{G}^*_{n+q} : \ N_n(\mathbf{i}) \geq 2\} \neq \varnothing, \ \mathcal{Z}_n \leq K\},$$

we have $$\mathbb{1}_{E^q_n} \leq \sum_{\mathbf{i}\in\mathbb{G}_{n+q}} \mathbb{1}_{\{N_n(\mathbf{i})\geq 2, \ \mathcal{Z}_n \leq K\}}.$$

Thus we can follow the proof of Lemma 8.6.4.

$$\begin{aligned}
\mathbb{P}^{n+q}(E^q_n) &\leq \sum_{\mathbf{i}\in\mathbb{G}_{n+q}} \frac{\mathbb{P}(N_n(\mathbf{i}) \geq 2, \ \mathcal{Z}_n \leq K)}{\mathbb{P}(\mathcal{Z}_{n+q} > 0)} \\
&\leq \frac{\sum_{\mathbf{i}\in\mathbb{G}_{n+q}} \mathbb{P}(N_n(\mathbf{i}) \geq 2 \ , \ Z_{\mathbf{i}|n} \leq K)}{U(2m)^{n+q}} \qquad \text{using (8.7)} \\
&\leq \frac{\mathbb{P}(0 < Z_n \leq K)2^{-q} \sum_{i\in\mathbb{G}_q} \mathbb{P}_K(N_0(\mathbf{i}) \geq 2)}{Um^{n+q}} \\
&\leq \frac{\binom{K}{2} 2^{-q} \sum_{i\in\mathbb{G}_q} \mathbb{P}(Z_{\mathbf{i}} > 0)^2}{Um^q} \qquad \text{using (8.26).}
\end{aligned}$$

Conclude with (8.28). □

Proof of Theorem 8.5.7. STEP 1 : We recall that $\mathcal{P}_n$ is the set of parasites in generation n, follow STEP 1 in the proof of Theorem 8.5.2 and use its Notation. Thus, we begin with proving that for every $\epsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$,

$$\mathbb{P}^{n+q}\big(\parallel (\#\{\mathbf{i} \in \mathbb{G}^*_{n+q} : Z_{\mathbf{i}} = k\})_{k\in\mathbb{N}} - (N_k(n,q))_{k\in\mathbb{N}} \parallel_1 \neq 0\big) \leq \epsilon,$$

where for all $n, q, k \geq 0$, $\quad N_k(n,q) := \sum_{\mathbf{p}\in\mathcal{P}(n)} Y^q_k(\mathbf{p}).$

First, by (8.10), there exist $K, q_0 \in \mathbb{N}$ such that for every $q \geq q_0$,

$$\lim_{n\to\infty} \mathbb{P}^{n+q}(\mathcal{Z}_n > K) \leq \epsilon. \tag{8.38}$$

By Lemma 8.7.1, there exists $q_1 \geq q_0$ such that for every $n \in \mathbb{N}$, we have

$$\mathbb{P}^{n+q_1}\big(\{\mathbf{i} \in \mathbb{G}^*_{n+q_1} : \ N_n(\mathbf{i}) \geq 2\} \neq \varnothing, \ \mathcal{Z}_n \leq K\big) \leq \epsilon. \tag{8.39}$$

And by (8.38), there exists $n_0 \geq 0$ such that for every $n \geq n_0$,

$$\mathbb{P}^{n+q_1}(\mathcal{Z}_n \geq K) \leq 2\epsilon.$$

Then

$$\mathbb{P}^{n+q_1}\big(\#\{\mathbf{i} \in \mathbb{G}^*_{n+q_1} :\ N_n(\mathbf{i}) \geq 2\} \neq 0\big) \leq 3\epsilon.$$

Moreover

$$\#\{\mathbf{i} \in \mathbb{G}^*_{n+q_1} : N_n(\mathbf{i}) \geq 2\} = 0 \ \Rightarrow\ (\#\{\mathbf{i} \in \mathbb{G}^*_{n+q_1} : Z_{\mathbf{i}} = k\})_{k\in\mathbb{N}} = (N_k(n, q_1))_{k\in\ \mathbb{N}}$$

Then for every $n \geq n_0$,

$$\mathbb{P}^{n+q_1}\big(\ \| \ (\#\{\mathbf{i} \in \mathbb{G}^*_{n+q_1} : Z_{\mathbf{i}} = k\})_{k\in\mathbb{N}} - (N_k(n, q_1))_{k\in\ \mathbb{N}} \ \|_1 \neq 0\big) \leq 3\epsilon.$$

STEP 2 : As $l^1(\mathbb{N})$ is separable, we can consider the distance d associated with the weak convergence of probabilities on $l^1(\mathbb{N})$. It is defined for any $\mathbb{P}_1$ and $\mathbb{P}_2$ probabilities by (see Theorem 6.2 chapter II in [78])

$$d(\mathbb{P}_1, \mathbb{P}_2) = \sup\big\{\big|\int f(w)\mathbb{P}_1(dw) - \int f(w)\mathbb{P}_2(dw)\big| : \|f\|_\infty \leq 1,\ \|f\|_{Lips} \leq 1\big\}$$

where

$$\|f\|_{Lips} = \sup\big\{\frac{\|f(x) - f(y)\|_1}{\|x - y\|_1} : x, y \in \mathbb{S}^1(\mathbb{N}), x \neq y\big\}.$$

We prove now that for every $l \geq 1$, there exist $n_0(l) \in \mathbb{N}$ and a measure $\mu(l)$ on $\mathbb{N}^*$ such that for every $n \geq n_0(l)$,

$$d\big(\mathbb{P}^n\big((\#\{\mathbf{i} \in \mathbb{G}^*_n : Z_{\mathbf{i}} = k\})_{k\in\mathbb{N}} \in .\big),\ \mu(l)\big) \leq 1/2^l. \tag{8.40}$$

For that purpose, let $l \in \mathbb{N}$. By STEP 1, choose $q, n_0\ \in \mathbb{N}$ such that

$$\begin{gathered}\forall n \geq n_0, \\ d\big(\mathbb{P}^{n+q}\big((\#\{\mathbf{i} \in \mathbb{G}^*_{n+q} : Z_{\mathbf{i}} = k\})_{k\in\mathbb{N}} \in .\big),\ \mathbb{P}^{n+q}\big((N_k(n, q))_{k\in\mathbb{N}} \in .\big)\big) \leq 1/2^{l+1}.\end{gathered} \tag{8.41}$$

Recall that $(Y^q_k(\mathbf{p}))_{k\in\mathbb{N}}(\mathbf{p} \in \mathcal{P}(n))$ is an iid sequence distributed as $(Y^q_k)_{k\in\mathbb{N}}$ and $\#\mathcal{P}(n) = \mathcal{Z}_n$. Thus, under $\mathbb{P}^{n+q}$, $N_k(n, q)$ is the sum of $\mathcal{Z}_n$ variables which are iid, distributed as Y^q_k and independent of $\mathcal{Z}_n$, conditionally on $\sum_{k\in\mathbb{N}} \sum_{\mathbf{p}\in\mathcal{P}(n)} Y^q_k(\mathbf{p}) > 0$.

Moreover $\mathbb{P}^{n+q}(\mathcal{Z}_n \in .)$ converges weakly as $n \to \infty$ to a probability ν (see (8.9)) and we denote by $\mathcal{N}$ a random variable with distribution ν and by $(Y^q_k(i))_{k\in\mathbb{N}}(i \in \mathbb{N})$ an iid sequence independent of $\mathcal{N}$ and distributed as $(Y^q_k)_{k\in\mathbb{N}}$. Then we have for n large enough,

$$d\big(\mathbb{P}^{n+q}\big((N_k(n, q))_{k\in\ \mathbb{N}} \in .\big),\ \mu(l)\big) \leq 1/2^l, \tag{8.42}$$

where $\mu(l)$ is the distribution of $\big(\sum_{1\leq i\leq\mathcal{N}} Y^q_k(i)\big)_{k\in\ \mathbb{N}}$ conditionally on $\sum_{k\in\mathbb{N}} \sum_{1\leq i\leq\mathcal{N}} Y^q_k(i) > 0$. Combining (8.41) and (8.42) gives (8.40).

CONCLUSION : As $l^1(\mathbb{N})$ is complete, the space of probabilities on $l^1(\mathbb{N})$ endowed with d is complete (see Theorem 6.5 chapter II in [78]), $(\mu(l))_{l\in\mathbb{N}}$ converges and we get the convergence of Theorem 8.5.7.
We now prove that $\mathbb{E}(\sum_{k\in\mathbb{N}^*} kN_k) < \infty$. For all $n, K > 0$, we have

$$\mathbb{E}(\sum_{k\geq K} k\#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} = k\}|\mathcal{Z}_n > 0) \leq \mathbb{E}(\mathcal{Z}_n \mathbb{1}_{\{\mathcal{Z}_n \geq K\}}|\mathcal{Z}_n > 0) \leq \frac{\mathbb{E}(\mathcal{Z}_n^2)}{\mathbb{P}(\mathcal{Z}_n > 0)K}$$

which converges uniformly to 0 as $K \to \infty$ using (8.11). Moreover Theorem 8.5.7 and $k\#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} = k\} \leq \mathcal{Z}_n$ ensure that

$$\lim_{n\to\infty} \mathbb{E}(\sum_{1\leq k\leq K} k\#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} = k\} \mid \mathcal{Z}_n > 0) = \mathbb{E}(\sum_{1\leq k\leq K} kN_k).$$

Thus we get the expected limit

$$\mathbb{E}(\sum_{k\in\mathbb{N}} k\#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} = k\} \mid \mathcal{Z}_n > 0) \stackrel{n\to\infty}{\longrightarrow} \mathbb{E}(\sum_{k\in\mathbb{N}^*} kN_k)$$

and recalling Section 2.1, we have also

$$E(\sum_{k\in\mathbb{N}^*} k\#\{\mathbf{i} \in \mathbb{G}_n^* : Z_{\mathbf{i}} = k\} \mid \mathcal{Z}_n > 0) = \mathbb{E}(\mathcal{Z}_n \mid \mathcal{Z}_n > 0) \stackrel{n\to\infty}{\longrightarrow} \mathcal{B}'(1) < \infty.$$

This completes the proof. □

The proofs of the corollaries follow those of the previous section.

8.8 Fractal properties for $(m_0, m_1) \in D_3$

We want to study the fractal property of the boundary $\partial\mathbb{T}$ of the tree of cells, which we endow with usual metric

$$d(\mathbf{u}, \mathbf{v}) = e^{-\mathbf{u}\wedge\mathbf{v}} \quad (\mathbf{u}, \mathbf{v} \in \partial\mathbb{T})$$

where $\mathbf{u} \wedge \mathbf{v} = \max\{n \in \mathbb{N} : \mathbf{i}|n = \mathbf{j}|n\}$. It is then a compact set and the topology induced by d is the topology spanned by balls $(B(\mathbf{i}) : \mathbf{i} \in \mathbb{T})$ defined by

$$B(\mathbf{i}) = \{\mathbf{u} \in \partial\mathbb{T} : \mathbf{u}|n = \mathbf{i}\}.$$

For every $\mathbf{i} \in \mathbb{T}$, we define the random variables $W(\mathbf{i})$ and $W(\mathbf{i})^*$ as the following limit

$$W(\mathbf{i}) \stackrel{\mathbb{P}}{=} \lim_{n\to\infty} \frac{\#\{\mathbf{j} \in \mathbb{G}_n : Z_{\mathbf{ij}} > 0\}}{(m_0 + m_1)^n}, \qquad W(\mathbf{i})^* = \lim_{n\to\infty} \frac{\sum_{\mathbf{j}\in\mathbb{G}_n} Z_{\mathbf{ij}}}{(m_0 + m_1)^n} \quad \text{a.s.} \tag{8.43}$$

The first limit which holds in probability is a consequence of separation of descendances of parasites (see Section 8.6.3).The second limit holds a.s. and is a consequence of the Kesten-Stigum Theorem applied to the number of parasites which is a BGW process. Moreover

$$W(\mathbf{i}) = \sum_{k=1}^{Z_{\mathbf{i}}} W_k, \qquad W(\mathbf{i})^* = \sum_{k=1}^{Z_{\mathbf{i}}} W_k^* \quad \text{a.s.} \tag{8.44}$$

where $(W_k)_{k\in\mathbb{N}}$ and $(W_k^*)_{k\in\mathbb{N}}$ are iid sequences which are independent of $(Z_{\mathbf{j}})_{\mathbf{j}\in\mathbb{T}:|\mathbf{j}|\leq|\mathbf{i}|}$ and distributed resp. as $W := W(\varnothing)$ and $W^* := W^*(\varnothing)$.

We introduce now the branching measure on $\partial\mathbb{T}$: $\mu^{(\mathtt{br})}$. For each $\mathbf{i} \in \mathbb{G}_n$, we let

$$\mu^{(\mathtt{br})}(B(\mathbf{i})) = (m_0 + m_1)^{-n} W(\mathbf{i}). \tag{8.45}$$

Recall also that Corollary 8.5.3 ensures the following identities :

$$\mathrm{Ext}^c = \{\partial\mathbb{T}^* \neq \varnothing\} = \{\mu^{(\mathtt{br})}(\partial\mathbb{T}) \neq 0\} \quad \text{a.s..}$$

Let us check that $\mu^{(\mathtt{br})}$ given by (8.45) gives a.s. a measure on $\partial\mathbb{T}$. Indeed, we have a.s.

$$\forall \mathbf{i} \in \mathbb{T} \ : \quad \mu^{(\mathtt{br})}(B(\mathbf{i}0))) + \mu^{(\mathtt{br})}(B(\mathbf{i}1)) = \mu^{(\mathtt{br})}(B(\mathbf{i}))$$

and the result follows from Caratheodory's Theorem applied to the π system $(B(\mathbf{i}) : \mathbf{i} \in \mathbb{T})$. Recalling that $(m_0, m_1) \in D_3$, we have the following result.

Proposition 8.8.1. *We have a.s. the following limit which holds for $\mu^{(\mathtt{br})}$ almost every $\mathbf{u} \in \partial\mathbb{T}^*$,*

$$\lim_{n\to\infty} -\log\big(\mu^{(\mathtt{br})}(B(\mathbf{u}|n)\big)/n = \log(m_0 + m_1).$$

We can then give the Hausdorff dimension of the boundary of the tree of contaminated cells :

Corollary 8.8.2. *Conditionally on Ext^c, we have a.s.*

$$\dim\big(\partial\mathbb{T}^*\big) = \log(m_0 + m_1).$$

One perspective is now to give the untypical values for the branching measures given by largely contaminated lines. This is related with the number of cells contaminated by an untypical number of parasites. This is a work in progress with Julien Berestycki and Amaury Lambert and requires to determine first the large deviations of Branching Processes in Random Environment.

Before the proofs, we need some technical results. Note that W is stochastically dominated by W^*, since the number of cells contaminated at generation $n+q$ by a parasite at genration n is less than the number of parasites at generation $n+q$ whose ancestor is this parasite. Moreover as $Z^{(0)}$ and $Z^{(1)}$ have finite second moment, we have (use [8] or (8.11))

$$\sup_{n\in\mathbb{N}} \left\{\mathbb{E}\big((\mathcal{Z}_n/m^n)^2\big)\right\} < \infty.$$

This entails that for every $\alpha \in [0,2[$,

$$\mathbb{E}(W^\alpha) \leq \mathbb{E}(W^{*\alpha}) < \infty.$$

Lemma 8.8.3. *There exists a constant $M > 0$ such that for all $n, q \in \mathbb{N}$,*

$$\mathbb{E}\big(\#\{\mathbf{i} \in \mathbb{G}^*_{n+q} : Z_{\mathbf{i}|n} > K\}\big) \leq M \frac{(m_0+m_1)^{n+q}}{K}.$$

Proof. Fubini ensures that

$$\begin{aligned}
\mathbb{E}\big(\#\{\mathbf{i} \in \mathbb{G}^*_{n+q} : Z_{\mathbf{i}|n} > K\}\big) &\leq \sum_{\mathbf{i}\in\mathbb{G}_{n+q}} \mathbb{P}(Z_{\mathbf{i}|n} > K, Z_{\mathbf{i}} > 0) \\
&\leq \sum_{k>K}\sum_{\mathbf{j}\in\mathbb{G}_n} \mathbb{P}(Z_{\mathbf{j}} = k) \sum_{\mathbf{i}\in\mathbb{G}_q} \mathbb{P}_k(Z_{\mathbf{i}} > 0)
\end{aligned}$$

Moreover $\mathbb{P}_k(Z_{\mathbf{i}} > 0) = 1 - (1-\mathbb{P}(Z_{\mathbf{i}} > 0))^k \leq k\mathbb{P}(Z_{\mathbf{i}} > 0)$ and we have

$$\begin{aligned}
\mathbb{E}\big(\#\{\mathbf{i} \in \mathbb{G}^*_{n+q} : Z_{\mathbf{i}|n} > K\}\big) &\leq 2^q\mathbb{P}(Z_q > 0) \sum_{k>K}\sum_{\mathbf{j}\in\mathbb{G}_n} k\mathbb{P}(Z_{\mathbf{j}} = k) \\
&\leq (2m)^q 2^n \mathbb{E}(Z_n \mathbb{1}_{\{Z_n > K\}}) \\
&\leq (2m)^q 2^n K^{-1} \mathbb{E}(Z_n(Z_n - 1)) \\
&\leq (2m)^q 2^n K^{-1} \frac{\widetilde{m}(m^n - \bar{m}^n)}{(m-\bar{m})} \qquad \text{using (8.11)} \\
&\leq (2m)^{n+q} \frac{\widetilde{m}}{K(m-\bar{m})} \qquad \text{since } \bar{m} < m
\end{aligned}$$

which completes the proof. □

Proof of Proposition 8.8.1. The proof is close from the proof of the analogous result for BGW tree (see e.g. [52]). Complications come from the fact that the number of contaminated daughters of a cell is strongly linked to the number of parasites of this cell. So the numbers of contaminated daughters of cells $\mathbf{i} \in \mathbb{T}^*$ (or $\mathbf{i} \in \mathbb{G}_n^*$) are neither independent nor identically distributed.

1) First note that for every $n \in \mathbb{N}$, a.s. for every $\mathbf{u} \in \partial\mathbb{T}^*$, we have $\mu^{(\mathtt{br})}(B(\mathbf{u}|n)) > 0$ and

$$\begin{aligned}
& \int_{\partial\mathbb{T}^*} \frac{1}{(m_0+m_1)^n \mu^{(\mathtt{br})}(B(\mathbf{u}|n))} \mu^{(\mathtt{br})}(d\mathbf{u}) \\
= & \sum_{\substack{\mathbf{i}\in\mathbb{G}_n \\ \exists \mathbf{u}\in\partial\mathbb{T}^*,\ \mathbf{u}|n=\mathbf{i}}} \frac{1}{(m_0+m_1)^n \mu^{(\mathtt{br})}(B(\mathbf{i}))} \mu^{(\mathtt{br})}(B(\mathbf{i})) \\
= & \frac{\#\{\mathbf{i}\in\mathbb{G}_n : \exists \mathbf{u}\in\partial\mathbb{T}^*,\ \mathbf{u}|n=\mathbf{i}\}}{(m_0+m_1)^n} \\
\leq & \frac{\#\mathbb{G}_n^*}{(m_0+m_1)^n}
\end{aligned}$$

By Fubini $\mathbb{E}(\#\mathbb{G}_n^*/(m_0+m_1)^n) = 2^n \mathbb{P}(Z_n > 0)/(m_0+m_1)^n$, which is less than 1 since $(m_0, m_1) \in D_3$. Then,

$$M' = \sup_{n\in\mathbb{N}} \mathbb{E}\left(\int_{\partial\mathbb{T}^*} \frac{1}{(m_0+m_1)^n \mu^{(\mathtt{br})}(B(\mathbf{u}|n))} \mu^{(\mathtt{br})}(d\mathbf{u}) \right) < \infty$$

Note also that

$$\begin{aligned}
& \mu^{(\mathtt{br})}\big(\mathbf{u} : (m_0+m_1)^n \mu^{(\mathtt{br})}(B(\mathbf{u}|n)) < n^{-4}\big) \\
= & \mu^{(\mathtt{br})}\Big(\mathbf{u} : \frac{1}{(m_0+m_1)^n \mu^{(\mathtt{br})}(B(\mathbf{u}|n))} > n^4\Big) \\
\leq & n^{-4} \int_{\partial\mathbb{T}^*} \frac{1}{(m_0+m_1)^n \mu^{(\mathtt{br})}(B(\mathbf{u}|n))} \mu^{(\mathtt{br})}(d\mathbf{u})
\end{aligned}$$

so that taking expectations

$$\mathbb{E}\big[\mu^{(\mathtt{br})}\big(\mathbf{u} : (m_0+m_1)^n \mu^{(\mathtt{br})}(B(\mathbf{u}|n)) < n^{-4}\big)\big] \leq M' n^{-4}.$$

Markov inequality yields

$$\mathbb{P}\big[\mu^{(\mathtt{br})}\big(\mathbf{u} : (m_0+m_1)^n \mu^{(\mathtt{br})}(B(\mathbf{u}|n)) < n^{-4}\big) \geq n^{-2}\big] \leq M' n^{-2},$$

and Borel-Cantelli lemma ensures that a.s. for n large enough,

$$\mu^{(\mathtt{br})}\big(\mathbf{u} : (m_0+m_1)^n \mu^{(\mathtt{br})}(B(\mathbf{u}|n)) < n^{-4}\big) < n^{-2}.$$

Then Borel-Cantelli entails that a.s., for $\mu^{(\mathtt{br})}$ every $\mathbf{u} \in \partial\mathbb{T}$, we have for n large enough

$$(m_0+m_1)^n \mu^{(\mathtt{br})}(B(\mathbf{u}|n)) \geq n^{-4}.$$

Thus a.s., for $\mu^{(\mathtt{br})}$ every $\mathbf{u} \in \partial\mathbb{T}$,

$$\liminf_{n\to\infty} \log(\mu^{(\mathtt{br})}(B(\mathbf{u}|n)))/n \geq -\log(m_0 + m_1).$$

2) Let's prove that $-\log(m_0 + m_1)$ is also the upperbound.

$$\begin{aligned} & \mathbb{E}\big[\mu^{(\mathtt{br})}\big(\mathbf{u} : \mu^{(\mathtt{br})}(B(\mathbf{u}|n)) \geq x/(m_0+m_1)^n\big)\big] \\ = & \sum_{\mathbf{i}\in\mathbb{G}_n} \mathbb{E}\big[\mu^{(\mathtt{br})}(B(\mathbf{i}))\mathbb{1}_{W(\mathbf{i})\geq x}\big] \\ = & \sum_{\mathbf{i}\in\mathbb{G}_n} \mathbb{E}\big[\mu^{(\mathtt{br})}(B(\mathbf{i}))\mathbb{1}_{0<Z_{\mathbf{i}}\leq K,\ W(\mathbf{i})\geq x}\big] + \mathbb{E}\big[\mu^{(\mathtt{br})}(B(\mathbf{i}))\mathbb{1}_{Z_{\mathbf{i}}>K,\ W(\mathbf{i})\geq x}\big] \end{aligned}$$

Using (8.45) and (8.44), we get for every $K_n > 0$

$$\leq \quad (m_0+m_1)^{-n}\Big[\sum_{\mathbf{i}\in\mathbb{G}_n} \mathbb{E}\big[\mathbb{1}_{Z_{\mathbf{i}}>0}\big]\mathbb{E}\big[\mathbb{1}_{\sum_{k=1}^K W_k\geq x}\sum_{k=1}^K W_k\big] + \sum_{\mathbf{i}\in\mathbb{G}_n}\mathbb{E}\big[\mathbb{1}_{Z_{\mathbf{i}}>K}W(\mathbf{i})\big]\Big]$$

Using (8.43), Fatou lemma give and Minkowski inequality :

$$\mathbb{E}\big(\sum_{k=1}^K W_k^K\big)^{1/K} \leq \sum_{k=1}^K \mathbb{E}\big(W_k^K\big)^{1/K},$$

we get

$$\begin{aligned} \leq & \quad \frac{\mathbb{E}\big[\#\mathbb{G}_n^*\big]}{(m_0+m_1)^n}\frac{\mathbb{E}\big[(\sum_{k=1}^K W_k)^{3/2})\big]}{\sqrt{x}} + \limsup_{q\to\infty}\frac{\mathbb{E}\big[\#\{\mathbf{i}\in\mathbb{G}_{n+q}^* : Z_{\mathbf{i}|n} > K\}\big]}{(m_0+m_1)^{n+q}} \\ \leq & \quad M''\frac{K^{3/2}}{\sqrt{x}} + \frac{M}{K} \end{aligned}$$

for some constants M'', M using $\mathbb{E}(W^{3/2}) < \infty$ and Lemma 8.8.3. Making $K_n = n^2$ in this inequality ensures that for every $\epsilon > 0$,

$$\mathbb{E}\big[\mu^{(\mathtt{br})}\big(\mathbf{u} : \mu^{(\mathtt{br})}(B(\mathbf{u}|n)) \geq (1+\epsilon)^n/(m_0+m_1)^n\big)\big]$$

is summable. Then a.s.

$$\mu^{(\mathtt{br})}\big(\mathbf{u} : \mu^{(\mathtt{br})}(B(\mathbf{u}|n)) \geq (1+\epsilon)^n/(m_0+m_1)^n\big)$$

is summable and Borel-Cantelli entails that for $\mu^{(\mathtt{br})}$ every $\mathbf{u} \in \partial\mathbb{T}$, we have for n large enough

$$\mu^{(\mathtt{br})}(B(\mathbf{u}|n)) \leq (1+\epsilon)^n/(m_0+m_1)^n.$$

Thus a.s., for $\mu^{(\mathtt{br})}$ every $\mathbf{u} \in \partial\mathbb{T}$,

$$\limsup_{n\to\infty} \log(\mu^{(\mathtt{br})}(B(\mathbf{u}|n)))/n \leq -\log(m_0+m_1) + \log(1+\epsilon).$$

which completes the proof letting $\epsilon \to 0$. $\square$

Proof of Corollary 8.8.2. This is a classical consequence of Proposition 8.8.1 (see e.g. [52]). We work conditionally on Ext^c.

First let $0 < \alpha < \log(m_0+m_1)$. As $\mu^{(\mathtt{br})}(\partial\mathbb{T}) \neq 0$ a.s., Proposition 8.8.1 enables us to find a.s. a compact $K \subset \partial\mathbb{T}^*$ and $N \geq 0$ such that $\mu^{(\mathtt{br})}(K) \neq 0$ and

$$\forall \mathbf{u} \in K,\ \forall n \geq N \ : \quad -\log\big(\mu^{(\mathtt{br})}(B(\mathbf{u}|n)\big) \geq n\alpha. \tag{8.46}$$

For any open cover $(A_j)_{j\in\mathbb{N}}$ of K, we denote by $\mathbf{i}_j$ the most recent common ancestor of $A_j \cap K$. This latter satisfies

$$e^{-|\mathbf{i}_j|} = \mathrm{diam}(A_j \cap K).$$

Assume now that for every $j \in \mathbb{N}$, $\mathrm{diam}(A_j) \leq e^{-N}$, then $|\mathbf{i}_j| \geq N$ and by (8.46),

$$\mu^{(\mathtt{br})}(A_j \cap K) \leq \mu^{(\mathtt{br})}(B(\mathbf{i}_j) \cap K) \leq e^{-|\mathbf{i}_j|\alpha} = \mathrm{diam}(A_j \cap K)^\alpha.$$

So

$$\sum_{j\in\mathbb{N}} \mathrm{diam}(A_j)^\alpha \geq \sum_{j\in\mathbb{N}} \mu^{(\mathtt{br})}(A_j \cap K) \geq \mu^{(\mathtt{br})}(K) > 0$$

which ensures $\dim(\partial\mathbb{T}^*) \geq \dim(K) \geq \alpha$. Letting $\alpha \to \log(m_0 + m_1)$ gives

$$\dim(\partial\mathbb{T}^*) \geq \log(m_0 + m_1).$$

Conversely, $\partial\mathbb{T}^*$ can be covered by the balls $(B(\mathbf{i}))_{\mathbf{i}\in\mathbb{G}_n^*}$ of radius e^{-n}. As

$$\#\mathbb{G}_n^*(e^{-n})^{\log(m_0+m_1)} = (m_0 + m_1)^{-n}\#\mathbb{G}_n^*$$

converges in probability as n tends to infinity to a finite random variable W, we get for every $\alpha > \log(m_0 + m_1)$,

$$\lim_{\delta\to 0} \inf\{\sum_{j\in\mathbb{N}} \mathrm{diam}(A_j)^\alpha \ : \ \cup_{j\in\mathbb{N}} A_j \supset \partial\mathbb{T}^*,\ \mathrm{diam}(A_j) \leq \delta\} = 0.$$

Letting $\alpha \to \log(m_0 + m_1)$ gives

$$\dim(\partial\mathbb{T}^*) \leq \log(m_0 + m_1),$$

which completes the proof. □

Chapter 9

Cell contamination and branching process in random environment with immigration

9.1 Introduction

We consider the following model for cell division with parasite infection and state dependent contamination. We start with one cell which divides into two daughter cells at every generation. Each cell behaves independently and at each generation,
(i) parasites multiply randomly inside the cell,
(ii) a random number of parasites contaminate any cell from outside the cell population,
(iii) each cell divides into two daughter cells and the offspring of each parasite is shared randomly into the two daughter cells.

It is convenient to distinguish a first daughter cell called 0 and a second one called 1. We denote by $\mathbb{T} = \cup_{n\in\mathbb{N}}\{0,1\}^n$ the binary genealogical tree of the cell population, by $\mathbb{G}_n$ the set of cells at generation n and by $Z_{\mathbf{i}}$ the number of parasites of cell $\mathbf{i} \in \mathbb{T}$.

First, we describe by a branching process the random multiplication and sharing of parasites in the cell, i.e. this branching process combines (i) and (iii). Second, we describe the random contamination (ii) by immigration. Finally, we consider both to fully describe the model.

I Parasite infection and cell division

For every cell, we choose randomly a mechanism for multiplication of the parasites inside and sharing of their offspring when the cell divides. This mechanism is independent and identically distributed for every cell.

In that purpose, let $\mathbf{f}$ be a random bivariate probability generating function (p.g.f) , i.e. $\mathbf{f}$ is a.s. the p.g.f of a pair of random variables taking values in $\mathbb{N}$. Let $(f_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ be a sequence of iid couple p.g.f distributed as $\mathbf{f}$, which gives for every cell $\mathbf{i}$ the reproduction law and sharing of the offspring of its parasites.
More precisely, for every $\mathbf{i} \in \mathbb{T}$, let $(Z_k^{(0)}(\mathbf{i}), Z_k^{(1)}(\mathbf{i}))_{k\in\mathbb{N}}$ be a sequence of r.v. such that conditionally on $\mathbf{f_i} = \mathbf{g}$, $(Z_k^{(0)}(\mathbf{i}), Z_k^{(1)}(\mathbf{i}))_{k\in\mathbb{N}}$ are iid with common couple p.g.f $\mathbf{g}$:

$$\forall \mathbf{i} \in \mathbb{T},\ \forall k \in \mathbb{N},\ \forall s,t \in [0,1], \quad \mathbb{E}(s^{Z_k^{(0)}(\mathbf{i})} t^{Z_k^{(1)}(\mathbf{i})} \mid \mathbf{f_i} = \mathbf{g}) = \mathbf{g}(s,t).$$

In each generation, each parasite k of cell $\mathbf{i}$ gives birth to $Z_k^{(0)}(\mathbf{i}) + Z_k^{(1)}(\mathbf{i})$ children, $Z_k^{(0)}(\mathbf{i})$ of which go into the first daughter cell and $Z_k^{(1)}(\mathbf{i})$ of which into the second one, when the cell divides. This is a more general model for parasite infection and cell division than the model studied in Chapter 8, where there was no random environment ($\mathbf{f}$ was deterministic) and the the total number of parasites was a Galton Watson process. See [59] for the original model in continuous time.

Our model includes also the two following natural models, with random binomial repartition of parasites. Let Z be a random variable in $\mathbb{N}$ and $(P_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ be iid random variable in $[0,1]$. At each generation, every parasite multiplies independently with the same reproduction law Z. Thus parasites follow a Galton Watson process. Moreover $P_{\mathbf{i}}$ gives the mean fraction of parasites of the cell $\mathbf{i}$ which goes into the first daughter cell when the cell divides. More precisely, conditionally on $P_{\mathbf{i}} = p$, every parasite chooses independently the first daughter cell with probability p (and the second one with probability $1-p$).
It contains also the following model. Every parasite gives birth independently to a random cluster of parasites of size Z and conditionally on $P_{\mathbf{i}} = p$, every cluster of parasite goes independently into the first cell with probability p (and into the second one with probability $1-p$).

We want to take into account unequal sharing of parasites and do not make any assumption about $\mathbf{f}$, since these unequal sharing have been observed experimentally (M. de Paepe, G. Paul and F. Taddei at TaMaRa's Laboratory (Hôpital Necker, Paris) have infected the bacteria *E. Coli* with a lysogen bacteriophage M13, see [88]) . In Section 9.6.1, we consider this model (i.e. there is no contamination) and we determine when the organism recovers, in the sense that the number of infected cells becomes negligible compared to the number of cells when the generation tends to infinity. Actually, in this case, for any reproduction rate of parasites, we can find a necessary and sufficient condition on sharing of their offspring so that the organism recovers a.s., which generalizes results of Section 8.3 to random environment.

II State dependent contamination At each generation, each cell is contaminated by a random number of parasites which multiply randomly and are shared randomly between the two daughter cells. This contamination depends only on whether the cell is already infected or not.

That is, conditionally given the number of parasites in the cells, the numbers of parasites which contaminate the cells are independent. This contamination is identically distributed for infected cells, idem for non-infected cells. Formally, if a cell $\mathbf{i}$ contains x parasites, the contamination brings $Y_x^{(0)}$ parasites to the first daughter cell of $\mathbf{i}$ and $Y_x^{(1)}$ to the second one, where

$$\forall x \geq 1, \quad Y_1 :\overset{d}{=} Y_x^{(0)} \overset{d}{=} Y_x^{(1)}, \qquad Y_0 :\overset{d}{=} Y_0^{(0)} \overset{d}{=} Y_0^{(1)}.$$

Moreover we assume that contamination satisfies

$$0 < \mathbb{P}(Y_0 = 0) < 1, \qquad 0 < \mathbb{P}(Y_1 = 0), \tag{9.1}$$

which means that a given cell is not contaminated a.s., and every non-infected cell may be contaminated with a positive probability.

This model contains the case when the contamination is independent of the number of parasites in the cell (Y_0 and Y_1 are identically distributed). It also takes into account the case when only non infected cells can be contaminated ($Y_1 = 0$ a.s.) and the case when infected cells are 'weaker' and parasites contaminate them easier ($Y_1 \geq Y_0$ a.s.). For biological and technical reasons, we dot make Y_x depend on $x \geq 1$.

<u>III Cell division with parasite infection and contamination</u> We describe now the whole model. We start with a single cell with k parasites and denote by $\mathbb{P}_k$ the associated probability. Unless otherwise specified, we assume $k = 0$.

For every cell $\mathbf{i} \in \mathbb{T}$, conditionally on $Z_{\mathbf{i}} = x$ and $\mathbf{f_i} = \mathbf{g}$, the numbers of parasites $(Z_{\mathbf{i}0}, Z_{\mathbf{i}1})$ of its two daughter cells is distributed as

$$\sum_{k=1}^{x} (Z_k^{(0)}(\mathbf{i}), Z_k^{(1)}(\mathbf{i})) + (Y_x^{(0)}(\mathbf{i}), Y_x^{(1)}(\mathbf{i})),$$

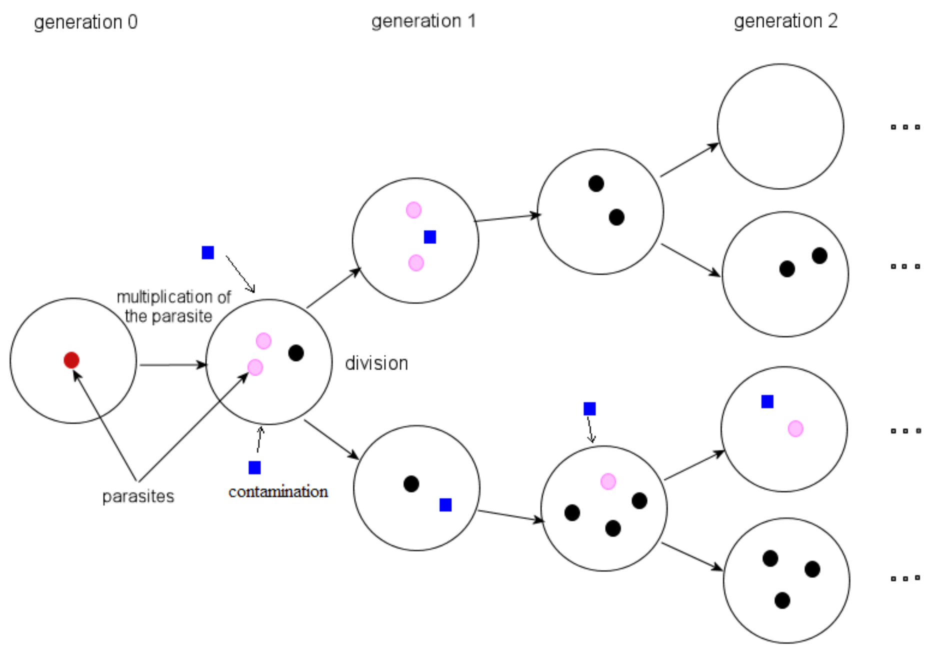
where

(i) $(Z_k^{(0)}(\mathbf{i}), Z_k^{(1)}(\mathbf{i}))_{k\geq 1}$ is an iid sequence with common couple probability generating function $\mathbf{g}$.

(ii) $(Y_x^{(0)}(\mathbf{i}), Y_x^{(1)}(\mathbf{i}))$ is independent of $(Z_k^{(0)}(\mathbf{i}), Z_k^{(1)}(\mathbf{i}))_{k\geq 1}$.

Moreover, $\big((Z_k^{(0)}(\mathbf{i}), Z_k^{(1)}(\mathbf{i}))_{k\geq 1},\ (Y_x^{(0)}(\mathbf{i}), Y_x^{(1)}(\mathbf{i}))_{x\geq 0}\big)$ are iid for $\mathbf{i} \in \mathbb{T}$.

Figure 8. Cell division with multiplication of parasites, random sharing and contamination. Each parasite gives bith to a random number of light parasites and dark parasites. Light parasites go into the first daughter cell, dark parasites go into the second daughter cell and square parasites contaminate the cell. But light/ dark/ square parasites then behave in the same way.



This model is a Markov chain indexed by a tree. This subject has been studied in the literature (see e.g. [9, 10, 18]) in the symmetric independent case. That is, $\forall(\mathbf{i}, k) \in \mathbb{T} \times \mathbb{N}$,

$$\mathbb{P}((Z_{\mathbf{i}0}, Z_{\mathbf{i}1}) = (k_0, k_1) \mid Z_{\mathbf{i}} = k) = \mathbb{P}(Z_{\mathbf{i}0} = k_0 \mid Z_{\mathbf{i}} = k)\mathbb{P}(Z_{\mathbf{i}0} = k_1 \mid Z_{\mathbf{i}} = k),$$

which would require that $Z^{(0)}$ and $Z^{(1)}$ are iid in this model. Guyon [47] proves limit theorems for a Markov chain indexed by a binary tree where asymmetry and dependence are allowed. His theorem is the key argument to prove convergence of asymptotic proportions of cells with a given number of parasites here. Indeed, contamination ensures that the random walk on the tree is ergodic and non trivial (see Section 9.5), which is the fundamental assumption to use his results.

Thus, a key role is played by the random walk on the binary tree, that is the number of parasites in a random cell line. More precisely, let $(a_i)_{i\in\mathbb{N}}$ be an iid sequence independent of $(Z_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ such that

$$\mathbb{P}(a_1 = 0) = \mathbb{P}(a_1 = 1) = 1/2. \tag{9.2}$$

Denote by $f^{(0)}$ (resp $f^{(1)}$) the random p.g.f which gives the law of the size of the offspring of a parasite which goes in the first daughter cell (resp. in the second daughter cell). That is,

$$f^{(0)}(s) = \mathbf{f}(s,1) \quad \text{a.s.}, \qquad f^{(1)}(t) = \mathbf{f}(1,t) \quad \text{a.s.}, \qquad (s,t \in [0,1]).$$

Let f be the mixed generating function of $f^{(0)}$ and $f^{(1)}$, i.e.

$$\mathbb{P}(f \in \mathrm{d}g) = \frac{\mathbb{P}(f^{(0)} \in \mathrm{d}g) + \mathbb{P}(f^{(1)} \in \mathrm{d}g)}{2}.$$

Then $(Z_n)_{n\in\mathbb{N}} = (Z_{(a_1,a_2,..a_n)})_{n\in\mathbb{N}}$ is a Branching Process in Random Environment with immigration depending on the state is zero or not : the reproduction law is given by its p.g.f f, the immigration law in zero is distributed as Y_0, and the immigration law in $k \geq 1$ as Y_1.

9.2 Main results

Galton Watson processes with immigration are well known (see e.g. [5, 70]). If the process is subcritical and the expectation of the logarithm of the immigration is finite, then it converges in distribution to a finite random variable. Otherwise it tends to infinity in probability. Key [58] has obtained the analogue result for Branching Processes in Random Environment with Immigration (IBPRE), in the subcritical case, with finite expectation of the logarithm. Actually he states results for multitype IBPRE, which have been complemented by Roitershtein who obtains a strong law of large numbers and a central limit theorem for the partial sum.

In Section 9.4, we give the asymptotic behavior of IBPRE in the general case, that is limit theorems for $(Z_n)_{n\in\mathbb{N}}$ in the case when Y_0 and Y_1 are identically distributed. To get this result, we begin with proving some results on Markov processes (Section 9.3.2), use classical arguments for Galton Watson process with immigration (see [70]) and the tail of the time when IBPRE returns to 0 in the subcritical case, which is proved in [58].

We can then state results about branching processes in random environment with immigration depending on the state is zero or not (Section 9.5) using coupling arguments and Section 9.3.2. This process gives the number of parasites along a random cell line. Recall that immigration in state zero is distributed as Y_0 and immigration in state $k \geq 1$ is distributed as Y_1. As expected, we prove that if $\mathbb{E}(\log(f'(1))) < 0$ and $\max(\mathbb{E}(\log^+(Y_i)) : i = 0,1) < \infty$, then this process converges in distribution. Otherwise, it tends to infinity in probability. More precisely

Theorem. *(i) If* $\mathbb{E}\big(\log(f'(1))\big) < 0$ *and* $\max(\mathbb{E}(\log^+(Y_i)) : i = 0,1) < \infty$*, then there exists a finite r.v.* Z_∞ *such that for every* $k \in \mathbb{N}$*,* Z_n *starting from* k *converges in distribution to* Z_∞ *as* $n \to \infty$*.*

(ii) If $\mathbb{E}\big(\log(f'(1))\big) \geq 0$ *or* $\max(\mathbb{E}(\log^+(Y_i)) : i = 0,1) = \infty$*,* Z_n *converges in probability to* ∞ *as* $n \to \infty$*.*

With additional assumptions, we provide in Section 9.5 an estimate of the rate of convergence of $(Z_n)_{n\in\mathbb{N}}$.

Then we prove asymptotics results on the population of cells in generation n as $n \to \infty$. First, we consider the case where there is no contamination : $Y_0 = Y_1 = 0$ a.s. and and we determine when the organism recovers, meaning that the number of contaminated cells N_n becomes negligible compared to the total number of cells (Section 9.6.1).

Proposition. *$N_n/2^n$ decreases as n grows.*
If $\mathbb{E}(\log(f'(1))) \leq 0$, then $N_n/2^n \to 0$ a.s. as $n \to \infty$.
Otherwise, $N_n/2^n \to 0$ as $n \to \infty$ iff all parasites die out, which happens with a probability less than 1.

In the case of the random binomial repartition of parasites with reproduction of parasites given by the r.v. Z and random sharing of parasites given by the r.v. $P \in [0,1]$ (see Introduction I), the organism recovers a.s. iff

$$\log(\mathbb{E}(Z)) \leq \mathbb{E}(\log(1/P)).$$

Then, we focus on proportions of cells in generation n with a given number of parasites in the general case (Section 9.6) :

$$F_k(n) := \frac{\#\{\mathbf{i} \in \mathbb{G}_n : Z_{\mathbf{i}} = k\}}{2^n} \qquad (k \in \mathbb{N}).$$

Using [47] and the theorem above, we prove that if $\mathbb{E}(\log(f'(1))) < 0$ and $\max(\mathbb{E}(\log^+(Y_i)) : i = 0,1) < \infty$, then proportions follow a law of large numbers. Otherwise, cells become largely infected. More precisely :

Theorem. *If $\mathbb{E}(\log(f'(1))) < 0$ and $\max(\mathbb{E}(\log^+(Y_i)) : i = 0,1) < \infty$, then for every $k \in \mathbb{N}$, $F_k(n)$ converges in probability to a deterministic number f_k as $n \to \infty$, such that $f_0 > 0$ and $\sum_{k=0}^{\infty} f_k = 1$.*
Otherwise, for every $k \in \mathbb{N}$, $F_k(n)$ converges in probability to 0 as $n \to \infty$.

Finally, we give the asymptotic behavior of the total number of parasites in generation n in the case when the growth of parasites follows a Galton Watson process and the contamination does not depend on the state of cell.

9.3 Preliminaries

We recall first some results about Branching Processes in Random Environment (BPRE) and then about Markov chains, which will be both useful to study BPRE with immigration $(Z_n)_{n\in\mathbb{N}}$. We denote by k the initial number of parasites and by $\mathbb{P}_k$ the probability associated with.

9.3.1 Branching Processes in Random Environment (BPRE)

We consider here a BPRE $(Z_n)_{n\in\mathbb{N}}$ specified by a sequence of iid generating functions $(f_n)_{n\in\mathbb{N}}$ distributed as f [6, 7, 87]. More precisely, conditionally on the environment $(f_n)_{n\in\mathbb{N}}$, particles at generation n reproduce independently of each other and their offspring has generating function f_n. Then Z_n is the number of particles at generation n and Z_{n+1} is the sum of Z_n independent random variables with generating function f_n. That is, for every $n\in\mathbb{N}$,

$$\mathbb{E}\big(s^{Z_{n+1}}|Z_0,\ldots,Z_n;\ f_0,\ldots,f_n\big)=f_n(s)^{Z_n}\qquad(0\le s\le 1).$$

Thus, denoting by $F_n:=f_0\circ\cdots\circ f_{n-1}$, we have for every $k\in\mathbb{N}$,

$$\mathbb{E}_k(s^{Z_{n+1}}\mid\ f_0,...,\ f_n)=\mathbb{E}(s^{Z_{n+1}}\mid Z_0=k,\ f_0,...,\ f_n)=F_n(s)^k\qquad(0\le s\le 1).$$

When the environments are deterministic (i.e. f is a deterministic generating function), this process is the Galton Watson process with reproduction law Z, where f is the generating function of Z .

The process $(Z_n)_{n\in\mathbb{N}}$ is called subcritical, critical or supercritical if

$$\mathbb{E}\big(\log(f'(1))\big)$$

is negative, zero or positive respectively. This process becomes extinct a.s. :

$$\mathbb{P}(\exists n\in\mathbb{N}:Z_n=0)=1$$

iff it is subcritical or critical [6] (see [43] for finer results).

In the critical case, we make the following integrability assumption :

$$0<\mathbb{E}(\log(f_0'(1))^2)<\infty,\quad \mathbb{E}\big([1+\log(f_0'(1))]f_0''(1)/2f_0'(1)\big)<\infty,$$

so that there exists $0<c_1<c_2<\infty$ such that for ever $n\in\mathbb{N}$ (see [63])

$$c_1/\sqrt{n}\le\mathbb{P}(Z_n>0)\le c_2/\sqrt{n}.\tag{9.3}$$

See [3] for more general result in the critical case.

9.3.2 Markov chains

We consider now a Markov chain $(Z_n)_{n\in\mathbb{N}}$ taking values in $\mathbb{N}$ and introduce T_0 the first time when $(Z_n)_{n\in\mathbb{N}}$ visits 0 after time 0 :

$$T_0:=\inf\{i>0:Z_i=0\}.$$

Denote by

$$u_n:=\mathbb{P}_0(Z_n=0),\qquad u_\infty:=1/\mathbb{E}_0(T_0)\qquad(1/\infty=0).$$

By now, we assume $\mathbb{P}_0(Z_1 = 0) > 0$ and we need and prove the following result which gives the asymptotic behavior of $(Z_n)_{n\in\mathbb{N}}$. The first part of (i) is the classical ergodicity of an aperiodic positive recurrent Markov chain and we provide an estimate of the convergence rate depending on the initial state. (ii) is the null recurrent case, which is also a classical result.

Lemma 9.3.1. *(i) If for every $k \in \mathbb{N}$, $\mathbb{E}_k(T_0) < \infty$, then Z_n starting from k converges in distribution to a finite random variable Z_∞, which does not depend on k and verifies*

$$\mathbb{P}(Z_\infty = 0) > 0.$$

Moreover there exists $A > 0$ such that for all $n, k \in \mathbb{N}$,

$$\begin{aligned}&\sum_{l\in\mathbb{N}} |\mathbb{P}_k(Z_n = l) - \mathbb{P}(Z_\infty = l)| \\ &\leq A\big[\sup_{n/2\leq l\leq n}\{|u_l - u_\infty|\} + \mathbb{E}_0(T_0 \mathbb{1}_{T_0>n/4}) + \mathbb{E}_k(T_0 \mathbb{1}_{T_0>n/4})\big].\end{aligned} \tag{9.4}$$

(ii) If $\mathbb{E}_0(T_0) = \infty$ and for every $l \in \mathbb{N}$, $\mathbb{P}_l(T_0 < \infty) > 0$, then for every $k \in \mathbb{N}$, $Z_n \to \infty$ in $\mathbb{P}_k$-probability as $n \to \infty$.

Proof of (i). As $\mathbb{P}_0(Z_1 = 0) > 0$, the renewal theorem [36] ensures that

$$u_n \overset{n\to\infty}{\longrightarrow} u_\infty.$$

First, note that by the Markov property,

$$\begin{aligned}&|\mathbb{P}_k(Z_n = 0) - u_\infty| \\ = \ & |\sum_{j=0}^{n} \mathbb{P}_k(T_0 = j)\mathbb{P}_0(Z_{n-j} = 0) - u_\infty| \\ \leq \ & \sum_{j=0}^{n} \mathbb{P}_k(T_0 = j)|u_{n-j} - u_\infty| + u_\infty \mathbb{P}_k(T_0 > n).\end{aligned} \tag{9.5}$$

On the event $\{T_0 \leq n\}$, define R_n as the last passage time of $(Z_n)_{n\in\mathbb{N}}$ by 0 before time n:

$$R_n := \sup\{i \leq n : Z_i = 0\}.$$

For all $0 \leq i \leq n$ and $l \in \mathbb{N}$, by the Markov property,

$$\mathbb{P}_k(Z_n = l) \ = \ \mathbb{P}_k(T_0 > n,\ Z_n = l) + \sum_{i=0}^{n} \mathbb{P}_k(T_0 \leq n,\ R_n = n - i,\ Z_n = l)$$

$$= \quad \mathbb{P}_k(T_0 > n,\ Z_n = l) + \sum_{i=0}^{n} \mathbb{P}_k(Z_{n-i} = 0)\mathbb{P}_0(Z_i = l,\ T_0 > i) \quad (9.6)$$

Define now

$$\alpha_l := u_\infty \sum_{i=0}^{\infty} \mathbb{P}_0(Z_i = l,\ T_0 > i).$$

We then have

$$\begin{aligned} |\mathbb{P}_k(Z_n = l) - \alpha_l| \quad &\leq \quad \mathbb{P}_k(T_0 > n,\ Z_n = l) + u_\infty \sum_{i=n+1}^{\infty} \mathbb{P}(Z_i = l,\ T_0 > i) \\ &\quad + \sum_{i=0}^{n} \mathbb{P}(Z_i = l,\ T_0 > i)\big|u_\infty - \mathbb{P}_k(Z_{n-i} = 0)\big|. \end{aligned}$$

Summing over l leads to

$$\begin{aligned} \sum_{l\in\mathbb{N}} |\mathbb{P}_k(Z_n = l) - \alpha_l| \quad &\leq \quad \mathbb{P}_k(T_0 > n) + u_\infty \mathbb{E}_0(T_0 1\!\!1_{T_0 > n+1}) \\ &\quad + \sum_{i=0}^{n} \mathbb{P}(T_0 > i)\big|u_\infty - \mathbb{P}_k(Z_{n-i} = 0)\big|. \end{aligned}$$

Moreover using (9.6), we have for all $0 \leq n_0 \leq n$,

$$\begin{aligned} &\sum_{i=0}^{n} \mathbb{P}(T_0 > i)\big|u_\infty - \mathbb{P}_k(Z_{n-i} = 0)\big| \\ \leq \quad &\sum_{i=0}^{n} \mathbb{P}_0(T_0 > i)[\sum_{j=0}^{n-i} \mathbb{P}_k(T_0 = j)|u_{n-i-j} - u_\infty| + u_\infty \mathbb{P}_k(T_0 > n-i)] \\ \leq \quad &\sum_{i=0}^{n} \mathbb{P}_0(T_0 > i)\sum_{j=0}^{n-i} \mathbb{P}_k(T_0 = j)|u_{n-i-j} - u_\infty| + u_\infty \sum_{i=0}^{n} \mathbb{P}_0(T_0 > i)\mathbb{P}_k(T_0 > n-i). \end{aligned}$$

Finally, denoting by $M := \sup_{n\in\mathbb{N}}\{|u_n - u_\infty|\}$,

$$\begin{aligned} &\sum_{i=0}^{n} \mathbb{P}_0(T_0 > i)\sum_{j=0}^{n-i} \mathbb{P}_k(T_0 = j)|u_{n-i-j} - u_\infty| \\ &\qquad \leq \sup_{n_0 \leq l \leq n}\{|u_l - u_\infty|\} \sum_{i=0}^{n} \mathbb{P}_0(T_0 > i)\sum_{j=0}^{n-i} \mathbb{P}_k(T_0 = j)1_{n-i-j \geq n_0} \\ &\qquad\quad + M\sum_{i=0}^{n} \mathbb{P}_0(T_0 > i)\sum_{j=0}^{n-i} \mathbb{P}_k(T_0 = j)1_{n-i-j < n_0} \\ &\qquad \leq \sup_{n_0 \leq l \leq n}\{|u_l - u_\infty|\} \sum_{i=0}^{n} \mathbb{P}_0(T_0 > i)\sum_{j=0}^{n-i} \mathbb{P}_k(T_0 = j) \end{aligned}$$

$$+M \sum_{i=0}^{n-n_0} \mathbb{P}_0(T_0 > i)\mathbb{P}_k(T_0 > n - n_0 - i).$$

Combining the three last inequalities and using that

$$\sum_{i=0}^{n} \mathbb{P}_0(T_0 > i)\mathbb{P}_k(T_0 > n - i) \leq \mathbb{E}_0(T_0 \mathbb{1}_{T_0 > n/2}) + \mathbb{E}_k(T_0 \mathbb{1}_{T_0 > n/2}),$$
$$\sum_{i=0}^{n-n_0} \mathbb{P}_0(T_0 > i)\mathbb{P}_k(T_0 \geq n - n_0 - i) \leq \mathbb{E}_0(T_0 \mathbb{1}_{T_0 > (n-n_0)/2}) + \mathbb{E}_k(T_0 \mathbb{1}_{T_0 > n - n_0)/2}),$$

we get, for all $0 \leq n_0 \leq n$,

$$\begin{aligned}\sum_{l \in \mathbb{N}} |\mathbb{P}_k(Z_n = l) - \alpha_l| \quad \leq \quad & \mathbb{P}_k(T_0 > n) + u_\infty \mathbb{E}_0(T_0 \mathbb{1}_{T_0 > n+1}) + \sup_{n_0 \leq l \leq n} \{|u_l - u_\infty|\} \mathbb{E}_0(T_0) \\ & + [u_\infty + M][\mathbb{E}_0(T_0 \mathbb{1}_{T_0 > (n-n_0)/2}) + \mathbb{E}_k(T_0 \mathbb{1}_{T_0 > (n-n_0)/2})].\end{aligned}$$

By renewal theorem $u_n \overset{n\to\infty}{\longrightarrow} u_\infty$. Adding that $\mathbb{E}_k(T_0) < \infty$ and $\mathbb{E}_0(T_0) < \infty$ ensures that

$$\sum_{l \in \mathbb{N}} |\mathbb{P}_k(Z_n = l) - \alpha_l| \overset{n\to\infty}{\longrightarrow} 0,$$

which proves that Z_n starting from k converges in distribution to a r.v. Z_∞ which does not depend on k.

The inequality of the lemma is obtained by letting $n_0 = n/2$. □

Proof of (ii). If $\mathbb{E}_0(T_0) = \infty$, then by the renewal theorem again [?],

$$u_n = \mathbb{P}(\exists k \in \mathbb{N} : T_k = n) \overset{n\to\infty}{\longrightarrow} 0.$$

So

$$D_n = \inf\{T_k - n : k \in \mathbb{N},\ T_k \geq n\} \overset{n\to\infty}{\longrightarrow} \infty, \qquad \text{in probability.}$$

Assume that there exists $l \in \mathbb{N}$, $\epsilon > 0$ and an increasing sequence of integers $(u_n)_{n \in \mathbb{N}}$ such that

$$\mathbb{P}_k(Z_{u_n} = l) \geq \epsilon.$$

As $\mathbb{P}_l(T_0 < \infty) > 0$ by hypothesis, there exists $N > 0$ such that

$$\mathbb{P}_l(T_0 = N) > 0.$$

Thus, by the Markov property,

$$\mathbb{P}_k(Z_{u_n+K} = 0) \geq \mathbb{P}_k(Z_{u_n} = l)\mathbb{P}_l(T_0 = N) \geq \epsilon \mathbb{P}_l(T_0 = N).$$

Then, for all $n \in \mathbb{N}$,

$$\mathbb{P}_k(D_{u_n} \leq N) \geq \epsilon \mathbb{P}_l(T_0 = N) > 0,$$

which is in contradiction with D_n tends to ∞ in $\mathbb{P}_k$ as $n \to \infty$. Then, $\mathbb{P}_k(Z_n = l) \to 0$ as $n \to \infty$. $\square$

9.4 Branching processes in random environment with immigration (IBPRE)

We consider here a BPRE $(Z_n)_{n\in\mathbb{N}}$ whose reproduction law is given by the random p.g.f f and we add at each generation n a random number of immigrants Y_n independent and identically distributed as a r.v Y such that

$$\mathbb{P}(Y = 0) > 0.$$

More precisely, for every $n \in \mathbb{N}$,

$$Z_{n+1} = Y_n + \sum_{i=1}^{Z_n} X_i, \tag{9.7}$$

where $(X_i)_{i\in\mathbb{N}}$, Y_n and Z_n are independent and conditionally on $f_n = g$, the $(X_i)_{i\in\mathbb{N}}$ are iid with common probability generating function g.

Note that if the contamination does not dependent on the fact that this cell is already contaminated or not (i.e. $Y_0 \stackrel{d}{=} Y_1$), then the number of parasites in a random cell line defined in Introduction is a IBPRE whose reproduction law given by f and immigration by $Y \stackrel{d}{=} Y_0 \stackrel{d}{=} Y_1$.

We give now the asymptotic behavior of this process. These results are classical for the Galton Watson process with immigration [70]. We follow the same method in the case of random environment for the subcritical and supercritical cases, give in (ii) the tail of the time when the process returns to 0 in the subcritical case, which is proved in [58] and we use Section 9.3.2 for the critical case.

Proposition 9.4.1. *(i) If $\mathbb{E}\big(log(f'(1))\big) < 0$ and $\mathbb{E}(\log^+(Y)) < \infty$, then Z_n converges in distribution to a finite random variable as $n \to \infty$ and* $\lim_{n\to\infty} \mathbb{P}(Z_n = 0) > 0$.
Otherwise $Z_n \to \infty$ as $n \to \infty$.

(ii) Recall that $T_0 := \inf\{n > 0 : Z_n = 0\}$*. If* $\mathbb{E}\big(log(f'(1))\big) < 0$ *and there exists* $q > 0$ *such that* $\mathbb{E}(Y^q) < \infty$*, then there exist* $c, d > 0$ *such that for every* $n \in \mathbb{N}$

$$\mathbb{P}(T_0 > n) \leq ce^{-dn}.$$

(iii) Assume $\mathbb{E}(f'(1)^{-1}) < 1$ *and* $\mathbb{E}(\log^+(Y)) < \infty$*, then there exists a finite r.v.* W *such that*

$$[\Pi_{i=0}^{n-1} f_i'(1)]^{-1} Z_n \overset{n\to\infty}{\longrightarrow} W, \quad \textit{in } \mathbb{P}.$$

Note that by the Borel-Cantelli lemma, if $\mathbb{E}(\log^+(Y_1)) = \infty$, then for every $c > 1$,

$$\limsup_{n\to\infty} c^{-n} Z_n = \infty \quad \text{a.s.}$$

since $Z_n \geq Y_n$ a.s. Moreover the proof of Section 9.5 provides an other approach to prove that $(Z_n)_{n\in\mathbb{N}}$ tends to ∞ if $\mathbb{E}(\log^+(Y)) = \infty$.

Proof of (i) and (ii) in the subcritical case : $\mathbb{E}\big(\log(f'(1))\big) < 0$. The subcritical case with assumption $\mathbb{E}(\log^+(Y)) < \infty$ is handled in [58] : First part of (i) is Theorem 3.3 and (ii) is a consequence of Theorem 4.2 of [58].

Assume that $\mathbb{E}\big(\log(f'(1))\big) < 0$ and $\mathbb{E}(\log^+(Y)) = \infty$. We prove now that Z_n converges in probability to ∞. The proof is close to the Galton Watson case (see [70]). First, by Borel-Cantelli lemma,

$$\limsup_{k\to\infty} \log^+(Y_k)/k = \infty \qquad a.s.$$

Then, for every $c \in]0, 1[$,

$$\limsup_{k\to\infty} c^k Y_k = \infty \qquad a.s. \tag{9.8}$$

Note that

$$Z_n = \sum_{k=0}^{n-1} Z_{k,n},$$

where $Z_{k,n}$ is the number of descendants in generation n of immigrants in generation $n-k$. Thus, denoting by $Y_{k,n}$ the number of immigrants in generation $n-k$ and $X_i(k,n)$ the number of descendants in generation n of immigrant i in generation $n-k$, we have

$$Z_n = \sum_{k=0}^{n-1} \sum_{i=1}^{Y_{k,n}} X_i(k,n).$$

This sum increases stochastically as n tends to infinity and tends in distribution to

$$Z_\infty = \sum_{k=0}^{\infty} \sum_{i=1}^{Y_k} X_i(k),$$

where conditionally on $(f_i : i \in \mathbb{N})$, $(X_i(k) \ : \ i \in \mathbb{N},\ k \in \mathbb{N})$ are independent and the probability generating function of $X_i(k)$ is equal to $f_{k-1} \circ ... \circ f_0$. Roughly speaking, $X_i(k)$ is the contribution of immigrant i which arrives k generations before 'final time' ∞. The integer $X_i(k)$ is the population in generation k of a BPRE without immigration starting from 1.

Assume now that $Z_\infty < \infty$ with a positive probability. As $(X_i(k) \ : \ k \in \mathbb{N},\ 1 \leq i \leq Y_k)$ are integers, then conditionally on $Z_\infty < \infty$, only a finite number of them are positive. Thus, by Borel-Cantelli lemma, conditionally on $(Z_\infty < \infty,\ Y_k : k \in \mathbb{N},\ f_i : i \in \mathbb{N})$,

$$\sum_{k=0}^{\infty} Y_k \mathbb{P}(X_1(k) > 0) < \infty \qquad \text{a.s.}$$

Moreover, by convexity, for all g p.g.f and $s \in [0,1]$,

$$\frac{1-g(s)}{1-s} = \frac{g(1)-g(s)}{1-s} \geq \frac{g(1)-g(0)}{1-0} = 1-g(0), \qquad (0 \leq s \leq 1).$$

Then $1-g(s) \geq (1-g(0))(1-s)$ and by induction, we have for every $k \in \mathbb{N}$,

$$\begin{aligned} \mathbb{P}(X_1(k) > 0 \mid f_i : i \in \mathbb{N}) &= 1 - f_{k-1} \circ ... \circ f_0(0) \\ &\geq \Pi_{i=0}^{k-1}(1 - f_i(0)) \\ &= \exp(S_k), \end{aligned}$$

where $S_k := \sum_{i=0}^{k-1} \log(1 - f_i(0))$. Thus, conditionally on $(Z_\infty < \infty,\ Y_k : k \in \mathbb{N},\ f_i : i \in \mathbb{N})$,

$$\sum_{k=0}^{\infty} Y_k \exp(S_k) < \infty \qquad \text{a.s.}$$

Thus, on the event $\{Z_\infty < \infty\}$ which has a positive probability, we get

$$\sum_{k=0}^{\infty} Y_k \exp(S_k) < \infty \text{ a.s.}$$

Moreover S_n is a random walk with negative drift $\mathbb{E}(\log(1-f_0(0)))$. So letting $\alpha < \mathbb{E}(\log(1-f_0(1)))$, $\mathbb{P}(S_n < \alpha n)$ decreases exponentially by classical large deviation results. Then by Borel-Cantelli lemma, S_n is less than αn for a finite number of n, and

$$L := \inf_{n \in \mathbb{N}} \{S_n - \alpha n\} > -\infty \qquad \text{a.s.}$$

Using that for every $k \in \mathbb{N}$, $S_k \geq \alpha k + L$ a.s., we get

$$\sum_{k=0}^{\infty} \exp(\alpha k) Y_k < \infty,$$

with positive probability. This is in contradiction with (9.8). Then $Z_\infty = \infty$ a.s. and Z_n converges in probability to ∞ as $n \to \infty$. $\square$

Proof of (i) in the critical and supercritical case : $\mathbb{E}\big(\log(f'(1))\big) \geq 0$. First, we focus on the critical case. Recall that $T_0 = \inf\{i > 0 : Z_i = 0\}$ and consider $(\bar{Z}_n)_{n\in\mathbb{N}}$ the BPRE associated with $(Z_n)_{n\in\mathbb{N}}$, that is the critical BPRE with reproduction law f and no immigration. Thanks to (9.3), there exists $c_1 > 0$ such that for ever $n \in \mathbb{N}$,

$$\mathbb{P}_1(\bar{Z}_n > 0) \geq c_1/\sqrt{n}.$$

Adding that

$$\mathbb{P}_1(T_0 > n) = \mathbb{P}_1(Z_n > 0) \geq \mathbb{P}_1(\bar{Z}_n > 0),$$

ensures that

$$\mathbb{E}_1(T_0) = \infty.$$

Then $\mathbb{E}_0(T_0) = \infty$ since IBPRE $(Z_n)_{n\in\mathbb{N}}$ starting from 1 is stochastically larger than $(Z_n)_{n\in\mathbb{N}}$ starting from 0. Moreover $\forall k \in \mathbb{N}$, $\mathbb{P}_k(T_0 < \infty) > 0$, since $\mathbb{P}_k(\bar{T}_0 < \infty) = 1$ and $\mathbb{P}(Y = 0) > 0$. Then Lemma 9.3.1 (ii) ensures that $Z_n \to \infty$ in $\mathbb{P}$ as $n \to \infty$.

For the supercrtical case, follow the proof in the critical case (or use the result with a coupling argument) to get that $Z_n \to \infty$ in probability as $n \to \infty$ $\square$

Proof of (iii). We follow [70] again. If $\mathbb{E}(\log^+(Y)) < \infty$, by Borel-Cantelli Lemma

$$\limsup_{k\to\infty} \log^+(Y_k)/k = 0.$$

Then for every $c > 1$,

$$\sum_{k=0}^{\infty} c^{-k} Y_k < \infty \qquad \text{a.s.} \tag{9.9}$$

Define

$$P_n := [\Pi_{i=0}^{n-1} f_i'(1)]^{-1},$$

and denote by $\mathcal{F}_n$ the σ-field generated by $(Z_i : 0 \leq i \leq n)$, $(P_i : 0 \leq i \leq n)$ and $(Y_k : k \in \mathbb{N})$. Then using (9.7), we have

$$\mathbb{E}(P_{n+1}Z_{n+1} \mid \mathcal{F}_n) \quad = \quad \mathbb{E}(P_{n+1}[\sum_{i=1}^{Z_n} X_i + Y_n] \mid \mathcal{F}_n)$$

$$\begin{aligned}
&= P_n\mathbb{E}(f_n'(1)^{-1}\sum_{i=1}^{Z_n} X_i \mid \mathcal{F}_n) + P_n\mathbb{E}(f'(1)^{-1})Y_n \\
&= P_n\mathbb{E}(f_n'(1)^{-1} Z_n \mathbb{E}(X_1 \mid f_n) \mid \mathcal{F}_n) + P_n\mathbb{E}(f'(1)^{-1})Y_n \\
&= P_n Z_n + P_n\mathbb{E}(f'(1)^{-1})Y_n.
\end{aligned}$$

So P_nZ_n is a submartingale. Moreover

$$\mathbb{E}(P_nZ_n \mid \mathcal{F}_0) \quad = \quad Z_0 + \sum_{i=0}^{n-1} \mathbb{E}(f'(1)^{-1})^{i+1} Y_i.$$

By (9.9), if $\mathbb{E}(f'(1)^{-1}) < 1$, P_nZ_n has bounded expectations and then converges a.s. to a finite r.v. □

9.5 Ergodicity and convergence for a random cell line

Recall that $(Z_n)_{n\in\mathbb{N}}$ defined in Introduction is the number of parasites in a random cell line. The Markov chain $(Z_n)_{n\in\mathbb{N}}$ is a BPRE with state dependent immigration. The reproduction law is given by the p.g.f f, immigration in state 0 is distributed as Y_0 and immigration in state $k \geq 1$ is distributed as Y_1. More precisely, for every $n \in \mathbb{N}$, conditionally on $Z_n = x$,

$$Z_{n+1} = Y_x^{(n)} + \sum_{i=1}^{x} X_i^{(n)},$$

where $(X_i^{(n)})_{i\in\mathbb{N}}$ and $Y_x^{(n)}$ are independent. Moreover conditionally on $f_n = g$, the $(X_i^{(n)})_{i\in\mathbb{N}}$ are iid with common probability generating function g. Recall also that for all $x \geq 1$ and $n \in \mathbb{N}$, $Y_x^{(n)} \stackrel{d}{=} Y_1$.

We have the following results, which generalize those of the previous section to the case when immigration depends on whether the state is zero or not.

Theorem 9.5.1. *(i) If* $\mathbb{E}\big(log(f'(1))\big) < 0$ *and* $\max(\mathbb{E}(\log^+(Y_i)) : i = 0, 1) < \infty$, *then there exists a finite random variable* Z_∞ *such that for every* $k \in \mathbb{N}$, Z_n *starting from* k *converges in distribution to* Z_∞ *as* $n \to \infty$.

Moreover, if there exists $q > 0$ *such that* $\max(\mathbb{E}(Y_i^q) : i = 0, 1) < \infty$, *then for every* $\epsilon > 0$, *there exist* $0 < r < 1$ *and* $C > 0$ *such that for all* $n \in \mathbb{N}$ *and* $k \in \mathbb{N}$,

$$\sum_{l=0}^{\infty} |\mathbb{P}_k(Z_n = l) - \mathbb{P}(Z_\infty = l)| \leq Ck^\epsilon r^n$$

(ii) If $\mathbb{E}\big(\log(f'(1))\big) \geq 0$ *or* $\max(\mathbb{E}(\log^+(Y_i)) : i = 0, 1) = \infty$*,* Z_n *converges in probability to infinity as* $n \to \infty$*.*

Note again that by Borel-Cantelli lemma, if $\mathbb{E}(\log^+(Y_1)) = \infty$, then for every $c > 1$,

$$\limsup_{n\to\infty} c^{-n} Z_n = \infty \quad \text{a.s.},$$

since $Z_n \geq Y_n$ a.s.

The proof of (ii) in the critical or supercritical case ($\mathbb{E}\big(\log(f'(1))\big) \geq 0$) is directly derived from Proposition 9.4.1 and we focus now on the subcritical case :

$$\mathbb{E}\big(\log(f'(1))\big) < 0.$$

Recall that T_0 is the first time after 0 when $(Z_n)_{n\in\mathbb{N}}$ visits 0. Using IBPRE (see Section 9.4), we prove the following result in the subcritical case.

Lemma 9.5.2. *If* $\max(\mathbb{E}(\log^+(Y_i)) : i = 0, 1) < \infty$*, then for all* $k, i \in \mathbb{N}$*,* $\mathbb{P}_k(T_i < \infty) = 1$ *and* $(Z_n)_{n\in\mathbb{N}}$ *is bounded in distribution :*

$$\sup_{n\in\mathbb{N}}\{\mathbb{P}_k(Z_n \geq l)\} \stackrel{l\to\infty}{\longrightarrow} 0.$$

Moreover if there exists $q > 0$ *such that* $\max(\mathbb{E}(Y_i^q) : i = 0, 1) < \infty$*, then for every* $\epsilon > 0$*, there exist* $r > 0$ *and* $C > 0$ *such that for all* $n \in \mathbb{N}$*,* $k \in \mathbb{N}$ *:*

$$\mathbb{P}_k(T_0 \geq n) \leq Ck^{\epsilon} r^n,$$
$$\mathbb{P}_0(T_0 \geq n) \leq Cr^n.$$

Proof. We couple $(Z_n)_{n\in\mathbb{N}}$ with an IBPRE $(\widetilde{Z}_n)_{n\in\mathbb{N}}$ with reproduction law given by the random p.g.f f (such as $(Z_n)_{n\in\mathbb{N}}$) and immigration Y defined by

$$Y := \max(Y_0,\ Y_1,\ \widetilde{Y}),$$

where $Y_0,\ Y_1$, and $\widetilde{Y}$ are independent and

$$\mathbb{P}(\widetilde{Y} = 0) = 1/2; \quad \forall n \in \mathbb{N}^*, \quad \mathbb{P}(\widetilde{Y} = n) = \alpha n^{-1-\epsilon}, \quad \alpha := [2\sum_{i=1}^{\infty} i^{-1-\epsilon}]^{-1}.$$

Thus immigration Y for $\widetilde{Z}_n$ is stochastically larger than immigration for Z_n (whereas reproduction law is the same), so that coupling gives

$$\forall n \in \mathbb{N}, \quad Z_n \leq \widetilde{Z}_n \qquad \text{a.s.}$$

Moreover, $\widetilde{Z}_n$ is still subcritical. Recalling that $\min(\mathbb{P}(Y_i = 0) : i = 0, 1) > 0$, $\mathbb{P}(\widetilde{Y} = 0) = 1/2$, and that the expectation of the logarithm of every r.v. is finite, we have

$$\mathbb{E}(\log^+(Y)) < \infty, \qquad \mathbb{P}(Y = 0) > 0.$$

Then Proposition 9.4.1 (i) ensures that $\widetilde{Z}_n$ converges in distribution to a finite random variable, so that

$$\sup_{n\in\mathbb{N}}\{\mathbb{P}_k(Z_n \geq l)\} \leq \sup_{n\in\mathbb{N}}\{\mathbb{P}_k(\widetilde{Z}_n \geq l)\} \overset{l\to\infty}{\longrightarrow} 0.$$

Proposition 9.4.1 (i) ensures also that $\lim_{n\to\infty}\mathbb{P}(Z_n = 0) > 0$. Thus, for all $k, i \in \mathbb{N}$, $\mathbb{P}_k(\widetilde{T}_i < \infty) = 1$, and then $\mathbb{P}_k(T_i < \infty) = 1$. This completes the first part of the lemma.

We assume now that there exists $q > 0$ such that $\max(\mathbb{E}(Y_i^q) : i = 0, 1) < \infty$. Moreover $\mathbb{E}(\widetilde{Y}^{\epsilon/2}) < \infty$, so letting $q' = \min(\epsilon/2, q)$, we have

$$\mathbb{E}(Y^{q'}) < \infty.$$

We can then apply Proposition 9.4.1 (ii) to IBPRE $(\widetilde{Z}_n)_{n\in\mathbb{N}}$, so that there exist $c, d > 0$ such that for every $n \in \mathbb{N}$,

$$\mathbb{P}_0(\widetilde{T_0} > n) \leq ce^{-dn}.$$

Note that for every $k \in \mathbb{N}$,

$$\mathbb{P}_0(\widetilde{T_0} > n) \geq \mathbb{P}(Y \geq k)\mathbb{P}_k(\widetilde{T_0} \geq n).$$

By definition of Y, there exists $\beta > 0$ such that for every $n \in \mathbb{N}$,

$$\mathbb{P}(Y \geq n) \geq \beta n^{-\epsilon},$$

Using the last three inequalities, we get

$$\begin{aligned}\mathbb{P}_k(T_0 \geq n) &\leq \mathbb{P}_k(\widetilde{T_0} \geq n)\\ &\leq \beta^{-1}k^{\epsilon}\mathbb{P}_0(\widetilde{T_0} > n)\\ &\leq \beta^{-1}ck^{\epsilon}e^{-dn}.\end{aligned}$$

This gives the first inequality of the lemma. Similarly

$$\mathbb{P}_k(T_1 - T_0 \geq n) \leq \mathbb{P}_k(\widetilde{T}_1 - \widetilde{T}_0 \geq n) = \mathbb{P}_0(\widetilde{T_0} \geq n) \leq ce^{-d(n-1)}.$$

This completes the proof of the lemma. □

Proof of Theorem 9.5.1 (i) and (ii) in the subcritical case : $\mathbb{E}\big(\log(f'(1))\big) < 0$.
We split the proof into 4 cases :

CASE 1 : $\max(\mathbb{E}(\log^+(Y_i)) : i = 0, 1) < \infty$.

CASE 2 : There exists $q > 0$ such that $\max(\mathbb{E}(Y_i^q) : i = 0, 1) < \infty$.

CASE 3 : $\mathbb{E}(\log^+(Y_1)) = \infty$.

CASE 4 : $\mathbb{E}(\log^+(Y_0)) = \infty$.

First, note that $\mathbb{P}(Y_0 = 0) > 0$ ensures that $\mathbb{P}_0(Z_1 = 0) > 0$ and we can use results of Section 9.3.2.

CASE 1. In this case, by Lemma 9.5.2, $(Z_n)_{n\in\mathbb{N}}$ is bounded in distribution :

$$\sup_{n\in\mathbb{N}}\{\mathbb{P}_k(Z_n \geq l)\} \stackrel{l\to\infty}{\longrightarrow} 0.$$

If $\mathbb{E}_0(T_0) = \infty$, then Z_n tends to ∞ in $\mathbb{P}_0$ by Lemma 9.3.1 (ii), which is in contradiction with the previous limit.
Then $\mathbb{E}_0(T_0) < \infty$. We prove now that $\forall k \geq 1$, $\mathbb{E}_k(T_0) < \infty$ by a coupling argument. Let $k \geq 1$ and change only immigration to get a Markov process $(\widetilde{Z}_n)_{n\in\mathbb{N}}$ which is larger than $(Z_n)_{n\in\mathbb{N}}$:

$$\forall n \in \mathbb{N}, \qquad \widetilde{Z}_n \geq Z_n \text{ a.s.}$$

Its immigrations $\widetilde{Y}_0$ and $\widetilde{Y}_1$ satisfy

$$\widetilde{Y}_1 \stackrel{d}{=} Y_1, \quad \forall n \in \mathbb{N},\ \mathbb{P}(\widetilde{Y}_0 \geq n) \geq \mathbb{P}(Y_0 \geq n),$$

$$\mathbb{P}(\widetilde{Y}_0 \geq k) > 0, \quad \max(\mathbb{E}(\log(\widetilde{Y}_i) : i = 0, 1) < \infty.$$

Then, we have again $\mathbb{E}_0(\widetilde{T}_0) < \infty$, which entails that $\mathbb{E}_k(\widetilde{T}_0) < \infty$ since $\mathbb{P}(\widetilde{Y}_0 \geq k) > 0$. As for every $n \in \mathbb{N}$, $\widetilde{Z}_n \geq Z_n$ a.s., we have

$$\mathbb{E}_k(T_0) \leq \mathbb{E}_k(\widetilde{T}_0) < \infty.$$

Then Lemma 9.3.1 (i) ensures that for every $k \in \mathbb{N}$, $(Z_n)_{n\in\mathbb{N}}$ converges in distribution to a finite random variable Z_∞, which does not depend on k and verifies $\mathbb{P}(Z_\infty = 0) > 0$.

CASE 2 : Recall that

$$T_0 := \inf\{i > 0 : Z_i = 0\}, \qquad T_{n+1} = \inf\{i > T_n : Z_i = 0\}.$$

and
$$u_n = \mathbb{P}(\exists k \in \mathbb{N} : T_k - T_0 = n), \qquad u_\infty := 1/\mathbb{E}_0(T_0).$$
By Lemma 9.3.1 (i), we have
$$\sum_{l\in\mathbb{N}} |\mathbb{P}_k(Z_n = l) - \mathbb{P}(Z_\infty = l)| \le A\big[\sup_{n/2\le l\le n}\{|u_l - u_\infty|\} + \mathbb{E}_0(T_0 \mathbb{1}_{T_0>n/4}) + \mathbb{E}_k(T_0\mathbb{1}_{T_0>n/4})\big]. \tag{9.10}$$
Moreover by Lemma 9.5.2, for every $\epsilon > 0$, there exists $C > 0$ such that
$$\mathbb{P}_k(T_0 \ge n) \le Ck^\epsilon r^n, \tag{9.11}$$
$$\mathbb{P}_0(T_0 \ge n) \le Cr^n. \tag{9.12}$$
So for every $r' \in]r, 1[$, $\mathbb{E}_0(\exp(-\log(r)T_0)) < \infty$. Then, by Kendall renewal theorem [57], there exists $\rho \in]0, 1[$ and $c > 0$ such that for every $n \in \mathbb{N}$,
$$|u_n - u_\infty| \le c\rho^n. \tag{9.13}$$
Finally, (9.11) and (9.12) ensure that there exists $D > 0$ such that for every $n \in \mathbb{N}$,
$$\begin{aligned} \mathbb{E}_0(T_0\mathbb{1}_{T_0>n/4}) &\le Dnr^{n/4}, \\ \mathbb{E}_k(T_0\mathbb{1}_{T_0>n/4}) &\le Dnk^\epsilon r^{n/4}. \end{aligned}$$
Combining these two inequalities with (9.10) and (9.13), we get
$$\sum_{l\in\mathbb{N}} |\mathbb{P}_k(Z_n = l) - \mathbb{P}(Z_\infty = l)| \le A\big[c\rho^n + Dnr^{n/4} + Dnk^\epsilon r^{n/4}\big],$$
which ends the proof in CASE 2.

CASE 3. Change immigration of $(Z_n)_{n\in\mathbb{N}}$ to get an IBPRE $(\widetilde{Z})_{n\in\mathbb{N}}$ whose immigration is distributed as Y_1 and whose reproduction law is still given by f. Then Proposition 9.4.1 (i) and $\mathbb{E}(\log^+(Y_1)) = \infty$ ensures that $(\widetilde{Z}_n)_{n\in\mathbb{N}}$ starting from 0 tends in distribution to ∞.
Then Lemma 9.3.1 (i) entails that $\mathbb{E}_0(\widetilde{T}_0) = \infty$, so that for every $k \ge 1$,
$$\mathbb{E}_k(\widetilde{T}_0) \ge \mathbb{E}_0(\widetilde{T}_0) = \infty,$$
since the IBPRE $(\widetilde{Z})_{n\in\mathbb{N}}$ starting from $k \ge 1$ is stochastically larger than $(\widetilde{Z})_{n\in\mathbb{N}}$ starting from 0.

Moreover, under $\mathbb{P}_k$, $(Z_n)_{n\in\mathbb{N}}$ is equal to $(\widetilde{Z}_n)_{n\in\mathbb{N}}$ until time $T_0 = \widetilde{T}_0$. So $\mathbb{E}_k(T_0) = \infty$. Let $k \ge 1$ such that $\mathbb{P}_0(Z_1 = k) > 0$, then $\mathbb{E}_0(T_0) \ge \mathbb{P}_0(Z_1 = k)\mathbb{E}_k(T_0 - 1)$. This entails that
$$\mathbb{E}_0(T_0) = \infty.$$

By Lemma 9.3.1 (ii), $(Z_n)_{n\in\mathbb{N}}$ starting from any $k \in \mathbb{N}$ tends to ∞ in probability.

CASE 4. Denote by

$$X_i := \mathbb{P}(Z_i > 0 \mid Z_{i-1} = 1,\ f_{i-1}), \quad (i \geq 1),$$

the survival probability in environment f_{i-1} and introduce the following random walk

$$S_n = \sum_{i=1}^{n} \log(X_i).$$

Then

$$\mathbb{P}_1(Z_n > 0 \mid (f_0, f_1, ..., f_{n-1})) \geq \Pi_1^n X_i = \exp(S_n) \quad \text{a.s.},$$

so that

$$\begin{aligned}
\mathbb{P}_k(Z_n > 0 \mid (f_0, f_1, ..., f_{n-1})) &= 1 - \mathbb{P}_k(Z_n = 0 \mid (f_0, f_1, ..., f_{n-1})) \\
&= 1 - [1 - \mathbb{P}_1(Z_n > 0 \mid (f_0, f_1, ..., f_{n-1}))]^k \\
&\geq 1 - [1 - \exp(S_n)]^k \quad \text{a.s.}
\end{aligned}$$

Thus

$$\mathbb{P}_k(Z_n > 0) \geq \mathbb{E}(1 - [1 - \exp(S_n)]^k).$$

Using the Markov property we have

$$\begin{aligned}
\mathbb{E}_0(T_0 + 1) &\geq \sum_{k=1}^{\infty} \mathbb{P}(Y_0 = k)\mathbb{E}_k(T_0) \\
&= \sum_{k=1}^{\infty} \mathbb{P}(Y_0 = k) \sum_{n=1}^{\infty} \mathbb{P}_k(T_0 \geq n) \\
&\geq \sum_{k=1}^{\infty} \mathbb{P}(Y_0 = k) \sum_{n=1}^{\infty} \mathbb{P}_k(Z_n > 0) \\
&\geq \sum_{k=1}^{\infty} \mathbb{P}(Y_0 = k) \sum_{n=1}^{\infty} \mathbb{E}(1 - [1 - \exp(S_n)]^k).
\end{aligned}$$

Moreover for all $x \in [0, 1[$ and $k \geq 0$, $\exp(k \log(1 - x)) \leq \exp(-kx)$, and by the law of large numbers, S_n/n tends a.s. to $\mathbb{E}(X_1) < 0$ so that there exists $n_0 \geq 1$ such that for every $n \geq n_0$,

$$\mathbb{P}(S_n/n \geq 3\mathbb{E}(X_1)/2) \geq 1/2.$$

We get then

$$\mathbb{E}_0(T_0 + 1) \geq \sum_{k=1}^{\infty} \mathbb{P}(Y_0 = k) \sum_{n=1}^{\infty} \mathbb{E}(1 - \exp(-k \exp(S_n)))$$

$$\begin{aligned}
&\geq [1-e^{-1}]\sum_{n=1}^{\infty}\sum_{k=1}^{\infty}\mathbb{P}(k\exp(S_n)\geq 1)\mathbb{P}(Y_0=k)\\
&\geq [1-e^{-1}]\sum_{n=n_0}^{\infty}\mathbb{P}(S_n/n\geq 3\mathbb{E}(X_1)/2)\sum_{k\geq \exp(-3n\mathbb{E}(X_1)/2)}^{\infty}\mathbb{P}(Y_0=k)\\
&\geq 2^{-1}[1-e^{-1}]\sum_{n=n_0}^{\infty}\mathbb{P}(Y_0\geq \exp(-3n\mathbb{E}(X_1)/2))\\
&\geq 2^{-1}[1-e^{-1}]\sum_{n=n_0}^{\infty}\mathbb{P}(\beta\log(Y_0)\geq n),
\end{aligned}$$

where $\beta := [-3\mathbb{E}(X_1)/2]^{-1} > 0$. Then $\mathbb{E}(\log(Y_0)) = \infty$ ensures that $\mathbb{E}_0(T_0+1) = \infty$, so

$$\mathbb{E}_0(T_0) = \infty.$$

Conclude that $(Z_n)_{n\in\mathbb{N}}$ tends to ∞ in $\mathbb{P}_k$ using Lemma 9.3.1 (ii). □

9.6 Asympotics for proportions of cells with a given number of parasites

9.6.1 Asymptotics without contamination

Here there is no contamination, i.e. $Y_0 = Y_1 = 0$ a.s. and we determine when the organism recovers, meaning that the number of contaminated cells becomes negligible compared to the total number of cells. We get the same result as Theorem 8.3.1 for the more general model considered here. Denote by N_n the number of contaminated cells.

Proposition 9.6.1. *$N_n/2^n$ decreases as n grows.*
If $\mathbb{E}(\log(f'(1))) \leq 0$, then $N_n/2^n \to 0$ a.s. as $n \to \infty$.
Otherwise, $N_n/2^n \to 0$ as $n \to \infty$ iff all parasites die out, which happens with a probability less than 1.

Example 8. Consider the case of the random binomial repartition of parasites (see Introduction). Let $Z \in \mathbb{N}$ be a r.v and $(P_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ be an iid sequence distributed as a r.v. $P \in [0,1]$, such that $P \stackrel{d}{=} 1-P$. In every generation, each parasite gives birth independently to a random number of parasites distributed as Z. When the cell $\mathbf{i}$ divides, conditionally on $P_{\mathbf{i}} = p$, each parasite of the cell $\mathbf{i}$ goes independently in the first daughter cell with probability p (or it goes in the second daughter cell, which happens with probability $1-p$). Then,

$$\mathbb{P}(f'(1) \in \mathrm{d}x) = \mathbb{P}(\mathbb{E}(Z)P \in \mathrm{d}x).$$

Thus, the organism recovers a.s. (i.e. $N_n/2^n$ tends a.s. to 0) iff

$$\log(\mathbb{E}(Z)) \leq \mathbb{E}(\log(1/P)).$$

This is the same criteria in the case when the offspring of each parasite goes a.s. is the same daughter cell (there, p is the probability that this offspring goes in the first daughter cell.)

In the case of non random environment for the cell (i.e. $\mathbf{f}$ is deterministic), which is the Kimmel branching model studied in [13], denoting by

$$m_0 := \partial\mathbf{f}(s,t)/\partial s(1,0) = \mathbb{E}(Z^{(0)}), \qquad m_1 := \partial\mathbf{f}(s,t)/\partial t(0,1) = \mathbb{E}(Z^{(1)}),$$

$\mathbb{E}(\log(f'(1))) < 0$ becomes

$$m_0 m_1 < 1.$$

Proof. Note that $N_n/2^n$ decreases to L as $n \to \infty$, since one contaminated cell has at most two daughter cells which are contaminated. Moreover, for every $n \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{E}\left(\frac{N_n}{2^n}\right) &= \frac{\mathbb{E}(\sum_{i\in\mathbb{G}_n} \mathbb{1}_{Z_\mathbf{i}>0})}{2^n} \\
&= \sum_{\mathbf{i}\in\mathbb{G}_n} \frac{1}{2^n}\mathbb{E}(\mathbb{1}_{Z_\mathbf{i}>0}) \\
&= \sum_{\mathbf{i}\in\mathbb{G}_n} \mathbb{P}((a_0, ..., a_{n-1}) = \mathbf{i})\mathbb{P}(Z_\mathbf{i} > 0) \\
&= \mathbb{P}(Z_n > 0).
\end{aligned}$$

If $\mathbb{E}(\log(f'(1))) \leq 0$ (subcritical or critical case for BPRE Z_n), then $\mathbb{P}(Z_n > 0)$ tends to 0 as $n \to \infty$ (see Section 9.3.1). Thus, $\mathbb{E}(L) = 0$ and $N_n/2^n$ tends to 0 a.s. as $n \to \infty$.

If $\mathbb{E}(\log(f'(1))) > 0$ (supercritical case for Z_n), then $\mathbb{P}(Z_n > 0)$ tends to a positive value p and $\mathbb{P}(L > 0) > 0$.

Let us prove that in the supercritical case, conditionally on non-extinction of parasites, the organism a.s. does recovers.

First, we prove that conditionally on non-extinction of parasites, for every $K \in \mathbb{N}$, there exists a generation n such that $N_n \geq K$. Let $K \in \mathbb{N}$, there exists $p \in \mathbb{N}$ such that $q := \mathbb{P}(N_p \geq K)q > 0$ since $\mathbb{P}(Z^{(0)} > 0,\ Z^{(1)} > 0) = \mathbb{P}(N_1 = 2) > 0$.

Either the number of contaminated cells in generation p is more than K (which happens with probability q), or we can choose in generation p a contaminated cell. Then, with probability larger than q , the number of contaminated cells in generation p of the subtree rooted in this cell contains more than K parasites. Note that this probability is equal to q if the contaminated cell which we have chosen contains one single parasite, as the first cell. Reasoning recursively, we find a generation with more than K contaminated cells.

Second, recalling that we still work conditionally on non-extinction of parasites, we can set $T = \inf\{n \in \mathbb{N} : N_n \geq K\} < \infty$ and work now conditionally on $T = n$ and $N_T = k$. Choose one parasite in every infected cell in generation n, which you label by $1 \leq i \leq k$. Denote then by $N_p^{(i)}$ the number of cells in generation $n+p$ infected by parasites whose ancestor in generation n is the parasite i. By branching property, The integers $(N_p(i) : 1 \leq i \leq k)$ are iid and $N_p^{(i)}/2^p \to L^{(i)}$ as $p \to \infty$, where $(L^{(i)} : 1 \leq i \leq k)$ are independent and $\mathbb{P}(L^{(i)} > 0) \geq \mathbb{P}(L > 0) > 0$. Using that

$$N_{n+p} = \sum_{i=1}^{k} N_p^{(i)} \qquad \text{a.s.},$$

we have, conditionally on $T = n$.

$$\lim_{p\to\infty} N_{n+p}/2^p \geq \max(L^{(i)} : 1 \leq i \leq K) \quad \text{a.s.}$$

As $\sup(L^{(i)} : i \in \mathbb{N}) = \infty$ a.s., letting $K \to \infty$ ensures that a.s. $N_p/2^p$ does not tend to 0. □

9.6.2 Asymptotics with contamination in the case $\mathbb{E}(\log(f'(1))) < 0$ and $\max(\mathbb{E}(\log^+(Y_i)) : i = 0, 1) < \infty$.

Define $F_k(n)$ the proportion of cells with k parasites in generation n :

$$F_k(n) := \frac{\#\{\mathbf{i} \in \mathbb{G}_n : Z_\mathbf{i} = k\}}{2^n} \qquad (k \in \mathbb{N}).$$

We introduce the Banach space $l^1(\mathbb{N})$ and the subset of frequencies $\mathbb{S}^1(\mathbb{N})$ which we endow with the norm $\| \,.\, \|_1$ defined by :

$$l^1(\mathbb{N}) := \{(x_i)_{i\in\mathbb{N}} : \sum_{i=0}^{\infty} |x_i| < \infty\}, \qquad \| (x_i)_{i\in\mathbb{N}} \|_1 = \sum_{i=0}^{\infty} |x_i|,$$

$$\mathbb{S}^1(\mathbb{N}) := \{(f_i)_{i\in\mathbb{N}} : \forall\, i \in \mathbb{N},\ f_i \in \mathbb{R}^+,\ \sum_{i=0}^{\infty} f_i = 1\}.$$

The main argument here is the law of large number proved by Guyon [47] for asymmetric Markov chains indexed by a tree.

Theorem 9.6.2. *If* $\mathbb{E}(\log(f'(1))) < 0$ *and* $\max(\mathbb{E}(\log^+(Y_i)) : i = 0, 1) < \infty$*, then* $(F_k(n))_{k\in\mathbb{N}}$ *converges in probability in* $\mathbb{S}^1(\mathbb{N})$ *to a deterministic sequence* $(f_k)_{k\in\mathbb{N}}$ *as* $n \to \infty$*, such that* $f_0 > 0$ *and* $\sum_{k=0}^{\infty} f_k = 1$*. Moreover, for every* $k \in \mathbb{N}$*,* $f_k = \mathbb{P}(Z_\infty = k)$*.*

Proof. Recall that $(Z_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ is a Markov chain indexed by a tree and we are in the framework of bifurcating Markov chain studied in [47]. Thanks to the ergodicity of the number of parasites in a random cell line proved in the previous section (Theorem 9.5.1 (i)), we can apply Theorem 8 in [47] to get the convergence of proportions of cells with a given number of parasites. □

But it seems that we can't apply Theorem 14 or Corollary 15 in [47] to get a.s. convergence of proportions, because of the term k^ϵ in estimation of Theorem 9.5.1.

In the case of the random binomial repartition of parasites given by $(P_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ and mutliplication of parasites given by $Z \in \mathbb{N}$, recall that criteria $\mathbb{E}(\log(f'(1))) < 0$ becomes

$$\log(\mathbb{E}(Z)) < \mathbb{E}(\log(1/P)).$$

Using again [47], we can prove also a law of large numbers and a central limit theorem for the proportions of cells with given number of parasites before generation n. Define, for every $n \in \mathbb{N}$,

$$P_k(n) := \frac{\#\{\mathbf{i} \in \cup_{0\le i\le n}\mathbb{G}_i : Z_{\mathbf{i}} = k\}}{2^{n+1}} \qquad (k \in \mathbb{N}).$$

Theorem 9.6.3. *If* $\mathbb{E}(\log(f'(1))) < 0$ *and* $\max(\mathbb{E}(\log^+(Y_i)) : i = 0, 1)$*, then* $(P_k(n))_{k\in\mathbb{N}}$ *converges in probability in* $\mathbb{S}^1(\mathbb{N})$ *to the deterministic sequence* $(f_k)_{k\in\mathbb{N}}$ *as* $n \to \infty$*.*
Moreover for every $k \in \mathbb{N}$*,* $\sqrt{n}(P_k(n) - f_k)$ *converges in distribution to a centered normal law as* $n \to \infty$*, with a non explicit variance.*

Proof. Use again Theorem 9.5.1 (i) and Theorem 8 in [47] to prove the law of large numbers. For the central limit theorem, use Theorem 19 in [47] by letting F be the set of measurable functions taking values in $[0, 1]$. □

9.6.3 Asymptotics with contamination in the case $\mathbb{E}(\log(f'(1))) \geq 0$ or $\max(\mathbb{E}(\log^+(Y_i)) : i = 0, 1) = \infty$.

In this case, cells become infinitely infected as the generation tends to infinity.

Theorem 9.6.4. *If $\mathbb{E}(\log(f'(1))) \geq 0$ or $\max(\mathbb{E}(\log^+(Y_i)) : i = 0, 1) = \infty$, for every $k \in \mathbb{N}$, $F_k(n)$ tends to zero as $n \to \infty$. That is, for very $K \in \mathbb{N}$, then*

$$\lim_{n\to\infty} \#\{\mathbf{i} \in \mathbb{G}_n : Z_{\mathbf{i}} \geq K\}/2^n \stackrel{\mathbb{P}}{=} 1.$$

Proof. By Fubini's theorem, we have

$$\begin{aligned}
\mathbb{E}\big[\#\{\mathbf{i} \in \mathbb{G}_n : Z_{\mathbf{i}} \geq K\}/2^n\big] &= \sum_{i\in\mathbb{G}_n} \mathbb{P}(Z_{\mathbf{i}} \geq K)/2^n \\
&= \sum_{i\in\mathbb{G}_n} \mathbb{P}((a_0, ...a_{n-1}) = \mathbf{i})\mathbb{P}(Z_{\mathbf{i}} \geq K) \\
&= \mathbb{P}(Z_n \geq K).
\end{aligned}$$

By Theorem 9.5.1, $\mathbb{P}(Z_n \geq K)$ tends to 1, then $1 - \#\{\mathbf{i} \in \mathbb{G}_n : Z_{\mathbf{i}} \geq K\}/2^n$ converges to 0 in L^1, which gives the result. □

9.7 Asymptotics for the number of parasites

We assume here that parasites multiply following a Galton Watson process with deterministic mean m, independently of the cell they belong to. That is, $s \mapsto \mathbf{f}(s, s)$ is deterministic and every parasite multiply independently with the reproduction law whose probability generating function is equal to $g : s \mapsto \mathbf{f}(s, s)$. Moreover we assume that contamination of a cell does not depend on the number of enclosed parasites. That is

$$Y \stackrel{d}{=} Y_0 \stackrel{d}{=} Y_1.$$

Set P_n the number of parasites in generation n. Without contamination, in the supercritical case $m > 1$, it is well know that either P_n becomes extinct or P_n/m^n converges to a positive finite random variable. In the precence of contamination, we have the following result.

Proposition 9.7.1. *If $\mathbb{E}(Y) < \infty$ and $\mathbb{P}(Y_0 > 0) > 0$, then $\log(P_n)/n$ converges in $\mathbb{P}$ to $\log(\max(2, m))$.*

Proof. First, we prove the lower bound. This is a direct consequence of the fact that P_n is larger than
(i) the number of parasites P_n^1 which contaminate one of the 2^n cells in generation n,

(ii) the number of parasites $P_n(p)$ in generation n with the same given ancestor in generation p.

First P_n^1 is the sum of 2^n iid random variables with mean $\mathbb{E}(Y)$, so law of large numbers ensures that

$$P_n^1/2^n \stackrel{n\to\infty}{\longrightarrow} \mathbb{E}(Y), \qquad \text{a.s.}$$

Moreover for every $p \in \mathbb{N}$,

$$P_n(p)/m^{n+p} \stackrel{n\to\infty}{\longrightarrow} W, \qquad \text{a.s.},$$

with $\mathbb{P}(W > 0) > 0$. Letting $p \to \infty$, the number of parasites in generation p tends to infinity a.s. (use P_p^1), so that for every $\epsilon > 0$, we can choose p such that the number P_n^2 of descendants of these parasites in generation n satisfies

$$P_n^2/m^{n+p} \stackrel{n\to\infty}{\longrightarrow} W', \qquad \text{a.s.}$$

with $\mathbb{P}(W' > 0) \geq 1 - \epsilon$. Using that N_n is larger than P_n^1 and P_n^2 ensures that for every $\epsilon > 0$,

$$\mathbb{P}(\log(P_n)/n \geq \log(\max(2,m)) - \epsilon) \stackrel{n\to\infty}{\longrightarrow} 1.$$

Second, we prove the upper bound. Note that the total number of parasites in generation n can be written as

$$P_n = \sum_{i=1}^{n} \sum_{j=1}^{2^i} \sum_{k=1}^{Y^{i,j}} Z_k^{i,j},$$

where $Y^{i,j}$ is the number of parasites which contaminate the jth cell of generation i, and labeling by $1 \leq k \leq Y^{i,j}$ these parasites, $Z_k^{i,j}$ is the number of descendants in generation n of the kth parasites.
Moreover $(Y^{i,j} : i \in \mathbb{N}, j \in \mathbb{N})$ are identically distributed and independent of $(Z_p^{i,j}(k), i \in \mathbb{N}, j \in \mathbb{N},\ k \in \mathbb{N})$, $(Z_k^{i,j}, i \in \mathbb{N}, j \in \mathbb{N},\ k \in \mathbb{N})$ are independent and $Z_p^{i,j}(k)$ is the population of a Galton Watson process in generation $n-i$ with offspring probability generation function equal to g. Thus

$$\begin{aligned}
\mathbb{E}(P_n) &= \sum_{i=1}^{n} \sum_{j=1}^{2^i} \mathbb{E}(\sum_{k=1}^{Y^{i,j}} Z_k^{i,j}) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{2^i} \mathbb{E}(Y^{i,j})\mathbb{E}(Z_k^{i,j}) \\
&= \mathbb{E}(Y) \sum_{i=1}^{n} \sum_{j=1}^{2^i} m^{n-i} \\
&= 2\mathbb{E}(Y)\frac{m^n - 2^n}{m-2} \quad \text{if } m \neq 2.
\end{aligned}$$

If $m = 2$, then $\mathbb{E}(P_n) = \mathbb{E}(Y)nm^n$. This gives the upper bound by Markov inequality and completes the proof. $\square$

Bibliography

[1] V. I. Afanasyev, J. Geiger, G. Kersting, V.A. Vatutin (2005). Functional limit theorems for strongly subcritical branching processes in random environment. *Stochastic Process. Appl.* 115, no. 10, 1658–1676.

[2] V. I. Afanasyev, J. Geiger, G. Kersting, V. A. Vatutin. Branching processes in random environment. A phase transition in the subcritical regime. Available via www.mathematik.uni-bielefeld.de/∼trutnau/kersting.pdf.

[3] V. I. Afanasyev, J. Geiger, G. Kersting, V. Vatutin (2005). Criticality for branching processes in random environment. *Ann. Probab.* **33** , no. 2, 645-673.

[4] A. Agresti (1975). On the extinction times of varying and random environment branching processes. *J. Appl. Prob.* 12, 39-46.

[5] S. Asmussen, H. Hering (1983). *Branching processes.* Progress in Probability and Statistics, 3. Birkhäuser Boston, Inc., Boston, MA.

[6] K. B. Athreya, S. Karlin (1971). On branching processes with random environments, I : Extinction probability. *Ann. Math. Stat.* 42. 1499-1520.

[7] K. B. Athreya, S. Karlin (1971). On branching processes with random environments, II : limit theorems. *Ann. Math. Stat.* 42. 1843-1858.

[8] K. B. Athreya, P. E. Ney (2004). *Branching processes.* Dover Publications, Inc., Mineola, NY.

[9] K. Athreya, H.J. Kang (1998). Some limit theorems for positive recurrent branching Markov chains I. *Adv. Appl. Prob.* 30(3). 693-710.

[10] K. Athreya, H.J. Kang (1998). Some limit theorems for positive recurrent branching Markov chains II. *Adv. Appl. Prob.* 30(3). 711-722.

[11] V. Bansaye (2007). On a model for the storage of files on a hardware I : Statistics at a fixed time and asymptotics. *Preprint* avialable via http://arxiv.org/abs/math/0611432.

[12] V. Bansaye (2007). On a model for the storage of files on a hardware II : Evolution of a typical data block. *J. Appl. Prob.* Vol 44, no 4, 901-927.

[13] V. Bansaye (2008). Proliferating parasites in dividing cells : Kimmel's branching model revisited. *Ann. Appl. Probab.*, Volume 18, Number 3, 967-996.

[14] V. Bansaye (2008). Limit theorems for subcritical branching processes in random environment depending on the initial number of particles. *Preprint* available via http://arxiv.org/abs/0804.0853.

[15] V. Bansaye (2008). Cell contamination and branching processes in random environment with immigration. *Preprint* available via http://arxiv.org/abs/0805.1863.

[16] A. D. Barbour, M. Kafetzaki (1993). A host-parasite model yielding heterogeneous parasite loads. *J. Math. Biol.* 31, no. 2, 157-176.

[17] A. D. Barbour, M. J. Luczak (2006). Laws of large numbers for epidemic models with countably many types. To appear in *Ann. Appl. Prob.*

[18] I. Benjamini, Y. Peres (1994). Markov chains indexed by trees. *Ann. Probab.* **22** (1). 219-243.

[19] J. Bertoin (1996). *Lévy processes.* Cambridge Tracts in Mathematics, 121. Cambridge University Press, Cambridge.

[20] J. Bertoin (1997). Subordinators: examples and applications. *Lectures on probability theory and statistics* (Saint-Flour), Lecture Notes in Math., 1717, Springer, Berlin, 1999.

[21] J. Bertoin (2000). Subordinators, Lévy processes with no negative jumps and branching processes. Lecture notes for MaPhySto, August 2000. Available via http://www.maphysto.dk/publications/MPS-LN/2000/8.pdf.

[22] J. Bertoin (2002). Eternal solutions to Smoluchowski's coagulation equation with additive kernel and their interpretation. *Annals of Applied Probability.* Vol. 12, no. 2, 547–564.

[23] J. Bertoin, G. Miermont (2006). Asymptotics in Knuth's parking problem for caravans. *Random Structures and Algorithms.* Vol.29, 38-55.

[24] P. Billingsley (1999). *Convergence of probability measures.* Second edition. Wiley Series in Probability.

[25] N. H. Bingham (1988). On the limit of a supercritical branching process. *J. Appl. Prob.* 25A, 215-228.

[26] N.H. Bingham, C.M. Goldie, J.L. Teugels (1989). *Regular variation.* Encyclopedia of Mathematics and its Applications, 27. Cambridge University Press, Cambridge.

[27] O. J. Boxma, J. W. Cohen (1999). Heavy-traffic analysis for the $GI/G/1$ queue with heavy-tailed distributions. *Queueing Systems Theory Appl. 33*, no. 1-3, 177–204.

[28] P. Chassaing, G. Louchard (2002). Phase transition for parking blocks, Brownian excursion and coalescence. *Random Structures Algorithms*. Vol. 21, no. 1, 76-119.

[29] P. Chassaing, P. Flajolet (2003). Hachage, arbres, chemins. *Gazette des mathématiciens*. no. 95, janvier 2003.

[30] J. W. Cohen (1982). *The single server queue*. Second edition. North-Holland Series in Applied Mathematics and Mechanics, 8. North-Holland Publishing Co., Amsterdam-New York.

[31] F. M. Dekking (1988). On the survival probability of a branching process in a finite state iid environment. *Stochastic Process. Appl.* 27, 151-157.

[32] S. Dubuc (1971). La densite de la loi limite d'un processus en casacade expansif. *Z. Wahr.* 19, 281-290.

[33] S. Dubuc (1971). Problemes relatifs a l'iteration des fonctions suggeres par les processus en cascade. *Ann. Inst. Fourier* 21, 171-251.

[34] R. Durrett (1980). Conditioned limit theorem for random walks with negative drift. *Z. Wahrsch. Verw. Gebiete 52*, no. 3, 277-287.

[35] Evans S. N., Steinsaltz D. (2006). Damage segregation at fissioning may increase growth rates : a superprocess model. Avialable via http://arxiv.org/PS_cache/q-bio/pdf/0608/0608008.pdf.

[36] W. Feller (1971). *An introduction to probability theory and its applications.* Vol. II. Second edition John Wiley and Sons, Inc., New York-London-Sydney.

[37] P. Flajolet, P. Poblete, A. Viola (1998). On the analysis of linear probing hashing. *Algorithica*. Vol. 22, no. 4, 490-515.

[38] K. Fleischmann, V. Vatutin (1999). Reduced subcritical Galton-Watson processes in a random environment. *Adv. in Appl. Probab.* 31, no. 1, 88-111.

[39] K. Fleischmann, V. Wachtel (2007). On the left tail asymptotics for the limit law of supercritical Galton-Watson processes in the Böttcher case. Avialable via http://www.wias-berlin.de/main/publications/wias-publ.

[40] P. J. Fitzsimmons, B. Fristedt, B. Maisonneuve (1985). Intersections and limits of regenerative sets. *Z. Wahrsch. Verw. Gebiete 70*, no. 2, 157–173.

[41] D. Foata, J. Riordan (1974). Mappings of acyclic and parking functions. *Aequationes Math. 10*, 10-22.

[42] J. Geiger (1999). Elementary new proofs of classical limit theorems for Galton-Watson processes. *J. Appl. Prob. 36*, 301-309.

[43] J. Geiger, G. Kersting, V. A. Vatutin (2003). Limit theorems for subcritical branching processes in random environment. *Ann. I. H. Poincaré* 39, no. 4. 593-620.

[44] P. Greenwood, J. Pitman (1980). Construction of local time and Poisson point processes from nested arrays. *J. London Math. Soc. (2) 22*, no. 1, 182-192.

[45] D.R. Grey, L. Zhunwei (1993). The asymptotic behaviour of extinction probability in the Smith Wilkinson branching processes. *Adv. Appl. Prob.* 25, No 2, 263-289.

[46] Y. Guivarc'h, Q. Liu (2001). Asymptotic properties of branching processes in random environment. *C.R. Acad. Sci. Paris,* t.332, Série I. 339-344.

[47] J. Guyon (2007). Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Ann. Appl. Probab.* **17**, no. 5-6, 1538-1569.

[48] P. Haccou, J. Jagers, V. A. Vatutin (2005). *Branching Processes: Variation, Growth and Extinction of Populations*. Cambridge University Press.

[49] B. Hambly (1992). On the limiting distribution of a supercritical branching process in random environment. *J. Appl. Probab.* 29, No 3, 499-518.

[50] K. Hirano (1998). Determination of the limiting coefficient for exponentials functional of random walks with positive drift. *J. Math. Sci. Univ. Tokyo 5,* 299-332.

[51] B. Hambly (1992). On the limiting distribution of a supercritical branching process in random environment. *J. Appl. Probab.* 29, No 3, 499-518.

[52] J. Hawkes (1981). Tree generated by a simple branching process. J. London Math. Soc. (2) 24, no. 2, 373–384.

[53] D. L. Iglehart (1974). Conditioned limit theorems for random walks. *Stochastic processes and related topics* (Proc. Summer Res. Inst. Statist. Inference for Stochastic Processes, Indiana Univ., Bloomington, Ind., 1974, Vol. 1; dedicated to Jerzy Neyman), 167-194.

[54] J. Jacod, A. N. Shiryaev (2003). *Limit theorems for stochastic processes.* Second edition. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 288. Springer-Verlag, Berlin.

[55] O. Kallenberg (2002). *Foundations of modern probability.* Second edition. Probability and its Applications (New York). Springer-Verlag, New York.

[56] H. Kaspi, M. Rubinovitch (1988). Regenerative sets and their applications to Markov storage systems. *Queueing theory and its applications*, 413–427, , CWI Monogr., 7, North-Holland, Amsterdam.

[57] D. Kendall (1959). Unitary dilations of Markov transition operators, and the corresponding integral representations for transition-probability matrices. *Probability and statistics* : The Harald Cramér volume (edited by Ulf Grenander.) 139-161.

[58] E. S. Key (1987). Limiting Distributions and Regeneration Times for Multitype Branching Processes with Immigration in a Random Environment. *Ann. Prob.*, Vol. 15, No. 1, 344-353.

[59] M. Kimmel (1997). Quasistationarity in a branching model of division-within-division. *Classical and modern branching processes (Minneapolis, MN, 1994)*. 157-164, IMA Vol. Math. Appl., 84, Springer, New York.

[60] M. Kimmel (2002). *Branching processes for biology.* Springer.

[61] J. F. C. Kingman (1965). The heavy traffic approximation in the theory of queues. *Proc. Sympos. Congestion Theory* (Chapel Hill, N.C., 1964) p. 137-169, Univ. North Carolina Press, Chapel Hill, N.C.

[62] J. F. C. Kingman (1993). Poisson processes. Oxford Studies in Probability, 3. Oxford Science Publications. *The Clarendon Press, Oxford University Press, New York*.

[63] M. V. Kozlov (1976). On the asymptotic behaviour of the probability of non-extinction for critical branching processes in a random environment. *Theory Probab. Appl.* 21 (4), 742-751.

[64] M. V. Kozlov (2006). On large deviations of branching processes in a random environment: a geometric distribution of the number of descendants. *Diskret. Mat. 18*, no. 2, 29-47.

[65] A. Lambert (2006). *Population Dynamics and Random Genealogies.* In Memorias del IX Simposio de probabilidad y procesos estocásticos (Guanajuato, México 2006). Edited by M.E. Caballero, V. Rivero, J. Ruiz de Chávez. To appear in *Stoch. Models.*

[66] Q. Liu (2000). On generalized multiplicative cascades. *Stochastic Process. Appl.* 86, no. 2, 263-286.

[67] C. J. Luchsinger (2001). Approximating the long-term behaviour of a model for parasitic infection. *J. Math. Biol.* 42, no. 6, 555-581.

[68] C. J. Luchsinger (2001). Stochastic models of a parasitic infection, exhibiting three basic reproduction ratios. *J. Math. Biol.* 42 (2001), no. 6, 532-554.

[69] R. Lyons, R. Pemantle, Y. Peres (1995). Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Ann. Probab.* 23 (3), 1125-1138.

[70] R. Lyons, Y. Peres (2005). *Probability on Trees and Networks.* Available via http://mypage.iu.edu/ rdlyons/prbtree/prbtree.html.

[71] B. Maisonneuve (1983). Ensembles régénératifs de la droite. [Regenerative sets of the line] *Z. Wahrsch. Verw. Gebiete 63*, no. 4, 501-510.

[72] G. Materon (1978). *Random sets and integral geometry.* J. Wiley & Sons, New York.

[73] P.A. Meyer (1971). Processus de Poisson ponctuels, d'après K. Itô. *Séminaire de Probabilités V.* Lectures Notes in Mathematics 191, Springer-Verlag, berlin, 209-210.

[74] P. W. Millar (1977). Zero-one laws and the minimum of a Markov process. *Trans. Am. Math. Soc.* Vol. 226, 365-391.

[75] P. Mörters, N.R. Shieh (2002). Thin and thick points for branching measure on a Galton Watson tree. *Stat. Prob. Letters* 58,13-22.

[76] P. Mörters, and M. Ortgiese (2008). Small value probabilities via the branching tree heuristic *Bernoulli,* Vol. 14, No 1, 277-299.

[77] P. E. Ney; A. N. Vidyashankar (2004). Local limit theory and large deviations for supercritical branching processes. *Ann. Appl. Probab.* 14, no. 3, 1135-1166.

[78] K. R. Parthasarathy (1967). *Probability measures on metric spaces.* Probability and Mathematical Statistics, No. 3 Academic Press, Inc., New York-London.

[79] J. Pitman (1986). Stationary excursions. In *Séminaire de Probabilités XXI*, Lecture notes in Math. 1247, 289-302. Springer.

[80] J. Pitman (2002). Combinatorial stochastic processes. Technical report n°621. *Lecture notes for Saint Flour course.* July 2002.

[81] N. U. Prabhu (1998). *Stochastic storage processes. Queues, insurance risk, dams, and data communication.* Second edition. Applications of Mathematics (New York), 15. Springer-Verlag, New York.

[82] P. Robert (2000). *Stochastic networks and queues.* Translated from the 2000 French edition. Applications of Mathematics (New York), 52. Stochastic Modelling and Applied Probability.

[83] A. Roitershtein (2007). A note on multitype branching processes with immigration in a random environment. *Ann. Probab.* 35, no. 4, 1573-1592.

[84] S. M. Ross (1983). *Stochastic Processes*. Wile, New York.

[85] A. Rouault (2000). Large deviations and branching processes. *Proceedings of the 9th International Summer School on Probability Theory and Mathematical Statistics (Sozopol, 1997).* Pliska Stud. Math. Bulgar. 13, 15–38.

[86] K. Sato (1999). Lévy processes and infinitely divisible distributions. Cambridge Studies in Advanced Mathematics, 68. *Cambridge University Press, Cambridge*.

[87] W.L. Smith, W. Wilkinson (1969). On branching processes in random environments. *Ann. Math. Stat.* 40, No 3, 814-827.

[88] E. J. Stewart, R. Madden, G. Paul, F. Taddei (2005). Aging and Death in a Organism that reproduces by Morphologically Symmetric Division. *PLoS Biol*, 3 (2) : e45.

[89] M. I. Taksar (1980). Regenerative sets on real line. *Séminaire de Probabilités XIV*, 437-474. Lecture Notes in Math., 784, Springer, Berlin.

[90] M. I. Taksar (1987). Stationary Markov sets. *Séminaire de Probabilités XXI*, 303-340. Lecture Notes in Math., 1247, Springer, Berlin.

[91] V. Vatutin (2004). A limit theorem for an intermediate subcritical branching process in a random environment. *Theory Probab. Appl. 48*, no. 3, 481-492.

[92] M. Winkel (2005). Electronic foreign-exchange markets and passage events of independent subordinators. *J. Appl. Probab.* Vol. 42, no. 1, 138–152.